

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I



TESIS DOCTORAL

**Estudio cristalográfico de los mecanismos de especificidad y
diseño molecular de glicosidasas para la producción de
oligosacáridos prebióticos y derivados bioactivos**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Mercedes Ramírez Escudero

Directora

Juliana Sanz Aparicio

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



TESIS DOCTORAL

**Estudio cristalográfico de los mecanismos de especificidad y
diseño molecular de glicosidasas para la producción de
oligosacáridos prebióticos y derivados bioactivos**

Memoria para optar al Grado de Doctor

Presentada por:

Mercedes Ramírez Escudero

Dirigida por:

Dra. Juliana Sanz Aparicio

DEPARTAMENTO DE CRISTALOGRAFÍA Y BIOLOGÍA ESTRUCTURAL

INSTITUTO DE QUÍMICA FÍSICA "ROCASOLANO"

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I

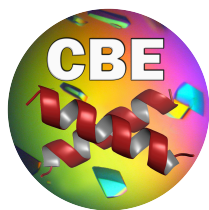


TESIS DOCTORAL

Estudio cristalográfico de los mecanismos de especificidad y diseño molecular de glicosidasas para la producción de oligosacáridos prebióticos y derivados bioactivos

Memoria presentada por Mercedes Ramírez Escudero para optar al Grado de Doctor

Esta Tesis Doctoral ha sido realizada en el Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano" (Consejo Superior de Investigaciones Científicas, CSIC) bajo la dirección de la Dra. Juliana Sanz Aparicio, Investigador Científico del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano", CSIC.



La Dra. Juliana Sanz Aparicio, Investigador Científico del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física “Rocasolano” del Consejo Superior de Investigaciones Científicas (CSIC) certifica que:

Doña Mercedes Ramírez Escudero, Licenciada en Biotecnología por la Universidad Francisco de Vitoria (Madrid) y Máster en Bioquímica, Biología Molecular y Biomedicina por la Universidad Complutense de Madrid, ha realizado bajo mi dirección el trabajo titulado “Estudio cristalográfico de los mecanismos de especificidad y diseño molecular de glicosidasas para la producción de oligosacáridos prebióticos y derivados bioactivos”, en el Departamento de Cristalografía y Biología Estructural del Instituto de Química Física “Rocasolano” (CSIC).

En mi opinión, este trabajo reúne las condiciones exigidas por la legislación vigente y tiene la originalidad, el rigor y la calidad científica necesarios y suficientes para ser presentado por Doña Mercedes Ramírez Escudero para optar al Grado de Doctor por la Universidad Complutense de Madrid.

Y para que así conste y con los efectos oportunos, firmo el presente certificado.

Madrid, ____ de _____ de 201__

Fdo.: Dra. Juliana Sanz Aparicio

Investigador Científico, CSIC

A mis padres

AGRADECIMIENTOS

Antes de comenzar con los aspectos técnicos y científicos, quiero agradecer a todas las personas que, de una manera u otra, han aportado su granito de arena y han hecho posible que esta tesis exista, y llegue a su fin. Espero no dejarme a nadie en el tintero.

En primer lugar, quisiera agradecer a la Dra. Julia Sanz Aparicio su apoyo y motivación durante estos cinco años de trabajo. Gracias por la confianza depositada en mí, por todo lo que me has enseñado y por ser una guía constante. También quiero agradecer a la Dra. Beatriz González su colaboración y ayuda con los protocolos de purificación y cristalización, así como por los constantes aportes en nuestras investigaciones, y por su amabilidad.

En segundo lugar, quiero agradecer a los colaboradores de estos trabajos de investigación. Gracias a la Dra. María Fernández Lobato, al Dr. Julio Polaina, al Dr. Manuel Ferrer y al Dr. Francisco Plou, y a los miembros de sus grupos de investigación, por la expresión y purificación de las proteínas, por los ligandos y por caracterizar y complementar los resultados cristalográficos tan eficientemente. En este apartado también quiero agradecer a otros colaboradores, aunque sus trabajos de investigación no aparezcan en estas páginas. Gracias al Dr. Javier Pastor, al Dr. Miguel Alcalde, a la Dra. Agatha Bastida y a la Dra. Esperanza Cerdán, así como a los miembros de sus grupos de investigación.

También quiero tener unas palabras de agradecimiento para los *local contact* de los sincrotrones, que tanta ayuda me han proporcionado durante mis numerosos viajes a Diamond, ALBA o ESRF. Especialmente, quiero agradecer al equipo de Xaloc del sincrotrón ALBA por su constante ayuda a la hora de realizar los experimentos de difracción, por mantener la línea en unas condiciones óptimas y atender a nuestras demandas con tanta amabilidad.

No puedo olvidarme de aquellos que me dieron mis primeras oportunidades. Gracias a la Dra. Maite Iglesias, porque además de directora y profesora, siempre supo guiar mis pasos en la Biotecnología. I want to thank Dr. Eishi Noguchi for giving me the opportunity to develop my interest in the scientific world, for being a guide and a motivation to become a PhD. También quiero dar las gracias al Dr. Miguel Díaz Hernández por acogerme en su laboratorio durante el master y permitirme aprender tanto de la neurociencia y los cultivos *in vitro*.

Además quiero agradecer al resto de Profesores, Investigadores, postdocs y predocs del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano" por sus aportaciones. De forma muy especial, quiero tener unas palabras para aquellos compañeros que se han convertido en personas muy queridas en mi vida. Gracias Iván por tus notas y por tus apuntes, chico 10, por tus risas dentro y fuera del labo. Gracias Javi por tu constante ayuda con la teoría cristalográfica, la fotografía y la informática, por tus correcciones y discusiones, a pesar de los cafés de máquina. Gracias María por esos viajes al sincro, no hubieran sido lo mismo sin ti. Gracias Mayte por haber traído frescura al departamento. Gracias Siseth por tu bondad y por aportar la nota "pícosa" al grupo. Gracias Martín y Juanlu por poner ese toque diferente y singular en el café. También quiero agradecer a otras grandes personas que han pasado por el Roca: Ángel, Agustín, Rocío, Ceci, Antonio, Nacho, Jorge, Almu, Ainhoa, Lisandro, Loreto, David, Patri, Anna, Erney y Ester.

En un apartado especial están los compañeros de despacho, que han hecho de ese espacio un pequeño hogar. Gracias Mangi por ayudarme en todo, por tus consejos, tus maravillosas

explicaciones, por allanarme el camino y facilitarme el trabajo. Gracias Yani por tu constante amor y simpatía, siempre con una palabra amable y una sonrisa, por compensar con “asúcar” los duros días de trabajo. Gracias Elsa por estar siempre ahí, por los viajes al sincro, a Granada, a Sevilla, ... que tanto hemos disfrutado, por estar siempre dispuesta a ayudar, y porque en las distancias cortas dejas de ser “cubito”. Gracias Noe por tu ternura, tu dulzura y tu amabilidad, por aportar ese arte que sólo tú tienes. Gracias Tere por poner diversión y locura en mi vida.

No puedo ni quiero olvidarme del resto de amigos que me han acompañado en todos estos años. A mis niñas del máster, Laura e Irene, que me han apoyado en todo este proceso, a veces nada fácil. También quiero agradecerle a mi compi Raquel por todos los años que hemos pasado juntas, por hacer de mami y cuidadora constante, incluso en la distancia. A mis biotech, que más de diez años después siguen siendo los mismos soñadores que aún no saben que van a ser de mayores, y a sus queridos y queridas, que ya son uno más del grupo. Muy especialmente a Lau, Zule y Reich, gracias por estar siempre ahí, sois tres pilares imprescindibles de mi vida.

También quiero darles las gracias a mis chiguitas, Bea, Mariola y Maribel, mis otros tres pilares, que aunque aún no entienden muy bien en qué consisten mis cristales y mis historias, me apoyan incondicionalmente. He crecido con vosotras, ¡y 18 años dan para muchas anécdotas, vivencias y fiestas! Gracias porque aunque no vivamos en la misma ciudad, parece que el tiempo no pasa cada vez que nos vemos. Muy especialmente quiero agradecer a Bea, mi hermana de otra madre, por estar siempre ahí. Contigo lo he pasado todo en la vida, lo bueno y lo malo. Gracias por entenderme y, más importante, por saber hacerme entrar en razón cuando es necesario.

Finalmente, quiero agradecer a mi familia, sin su apoyo no hubiera podido terminar esta tesis. Gracias a Tania, Cristina, Oscar, David, Sonia, Jose, Mila, Félix, Pili y Vitorino por vuestra ayuda y preocupación constante, por todos los momentos vividos. Especialmente a mi “pipi”, que ha sido muy paciente con la distancia y el tiempo. Gracias a mis abuelos, por cuidarme siempre con tanto cariño y hacerme una niña tan feliz.

Y como lo mejor lo dejo siempre para el final, gracias a mis padres, sin vosotros no sería nada. Gracias por vuestro cariño y amor, por vuestro constante ánimo e inmensa paciencia, por no dejarme caer ni perder el norte. Esta tesis es para vosotros, aunque es un pequeño reconocimiento comparado con todo lo que habéis hecho y hacéis por mí.

Según diversos estudios epidemiológicos, la alimentación juega un papel determinante en la salud humana. Se ha demostrado que la dieta tiene un efecto directo sobre la composición y el equilibrio de la microbiota intestinal y, ésta, sobre la salud del individuo. Por ello, se buscan nuevos alimentos “funcionales”, como los oligosacáridos prebióticos, que son capaces de estimular el crecimiento de bacterias beneficiosas.

Por otra parte, la glicosilación de distintos compuestos químicos mejora sus propiedades y su biodisponibilidad. Estos nuevos compuestos bioactivos aún no están tan desarrollados como los prebióticos, pero tienen una gran proyección de futuro en la industria farmacéutica.

La investigación sobre la biosíntesis de estos productos conlleva tanto el estudio de nuevos compuestos con mejores propiedades funcionales como la búsqueda de formas más eficaces de producirlos. En este aspecto, la Biotecnología Enzimática se ha impuesto frente a otras metodologías de química tradicional gracias a su excelente regio-, quimio- y estereo-especificidad. Y para profundizar en esta metodología, la Cristalografía de rayos X genera la información estructural básica para el diseño racional en la ingeniería de proteínas.

Por todo ello, esta tesis se ha centrado en el estudio cristalográfico de glicosil hidrolasas (GH) que son las enzimas capaces de metabolizar carbohidratos, los biocompuestos estructuralmente más diversos en la naturaleza. Estas enzimas, altamente específicas y eficientes, intervienen en multitud de procesos biológicos esenciales por lo que un conocimiento detallado de su función a nivel molecular es clave para entender dichos procesos y controlar numerosas enfermedades, además de permitir optimizar su aplicación biotecnológica. En este manuscrito, se presentan cuatro trabajos de investigación realizados con enzimas productoras de carbohidratos prebióticos del tipo FOS (fructooligosacáridos), IMOS (isomaltooligosacáridos) o β -glucanos, además de explorar su capacidad para sintetizar otros compuestos bioactivos.

El objetivo general de la investigación es avanzar en el esclarecimiento de los mecanismos de reconocimiento proteína-carbohidrato que modulan la especificidad de glicosidasas, mediante la Cristalografía de proteínas. Además de las implicaciones fundamentales para entender los procesos en los que estas enzimas intervienen, este conocimiento estructural es la base fundamental para realizar la ingeniería molecular de enzimas. El fin último es producir oligosacáridos con propiedades prebióticas, diseñados específicamente para manipular la microbiota intestinal del individuo, lo que abre la puerta a una futura medicina preventiva personalizada.

Las primeras enzimas estudiadas en esta tesis son la invertasa de *Saccharomyces cerevisiae* (ScInv) y la β -fructofuranosidasa de *Xanthophyllomyces dendrorhous* (XdInv), ambas de la familia GH32. Estas enzimas son bimodulares y están formadas por un dominio catalítico β -propeller y un dominio β -sándwich adyacente. Nuestros estudios cristalográficos han revelado que el dominio β -sándwich es responsable de la oligomerización y que este ensamblaje molecular regula la especificidad de sustrato de las enzimas GH32 de levadura. Mientras que ScInv es un octámero que restringe el acceso a sustratos largos, explicando su actividad esencialmente invertasa, XdInv presenta una organización del dímero distinto a la previamente descrita en la familia y que le podría permitir degradar fructanos mixtos ramificados. Esta propiedad podría estar relacionado con su función biológica en el hábitat natural de la levadura.

ScInv y XdInv son capaces de generar oligosacáridos prebióticos del tipo β (2-6)-FOS, con mayores capacidades prebióticas que los comercializados β (2-1)-FOS. Así, ScInv produce 6-kestosa y XdInv neokestosa. Un estudio en detalle del sitio activo de XdInv mediante el análisis de complejos

con diversos sustratos mostró las bases moleculares de su amplia especificidad y reveló la existencia de dos posibles sitios de unión alternativos, dependiendo del tipo de sustrato. En base a estos resultados, investigamos la posibilidad de utilizar como sustrato aceptor otros oligosacáridos, encontrando que la enzima es capaz de fructosilar maltooligosacáridos y producir hetero-FOS, con potencial capacidad prebiótica. Así mismo, hemos investigado la posibilidad de fructosilar otros compuestos orgánicos no-azúcares, generando nuevos compuestos bioactivos de interés como antioxidantes.

La α -glucosidasa de *Xanthophyllomyces dendrorhous* (Xd- α Glu) es una enzima con capacidad de producir IMOS del tipo $\alpha(1-6)$ de longitud media, con mejores propiedades funcionales y prebióticas. El modelo estructural por homología de secuencia revela que presenta una topología típica de las enzimas GH31 de eucariota. Sin embargo, presenta dos características estructurales novedosas: un inserto en el dominio N-terminal, y un Inserto 2 inusualmente largo en el dominio catalítico. Estas dos nuevas propiedades pueden ser las responsables de su preferencia por sustratos largos.

Por último, hemos estudiado una β -glucosidasa aislada del metagenoma de rumen bovino (GlyA₁). La estructura de GlyA₁ presenta una permutación de dominios, aunque conserva la arquitectura general de las β -glucosidasas de la familia GH3. Hemos demostrado que esta nueva topología está asociada a microorganismos que habitan en el tracto digestivo y que dependen de una dieta rica en biopolímeros de material vegetal. Además, GlyA₁ tiene un dominio extra C-terminal. Los análisis de dispersión de rayos X a bajo ángulo (SAXS) muestran una estructura compacta donde el dominio extra C-terminal se posiciona sobre el núcleo estructural de la enzima.

GlyA₁ es una β -glucosidasa/ β -xilosidasa de amplia especificidad, que presenta además otras actividades asociadas. Esta enzima es capaz de degradar oligosacáridos con distinto tipo de enlace y grado de polimerización, siendo más activa con sustratos de tamaño medio. Estas características la convierten en una buena candidata para nuevas aplicaciones biotecnológicas. Su mecanismo de retención y su amplia especificidad sugieren su potencial uso en la obtención de distintos prebióticos y oligosacáridos de interés biológico.

Para concluir, la Cristalografía de proteínas es una herramienta potente para determinar en detalle el mecanismo de acción de las enzimas. En esta tesis se presenta el estudio estructural de tres enzimas glicosil hidrolasa (ScInv, XdInv y GlyA₁) y de sus complejos, además de un modelo computacional de otra GH (Xd- α Glu). Nuestros resultados han permitido profundizar en el complejo mecanismo molecular que presenta cada enzima para reconocer los sustratos de forma específica. Este conocimiento de las proteínas a nivel atómico permite diseñar, de forma racional, enzimas con capacidades sintéticas mejoradas dirigidas a la producción de carbohidratos con capacidades prebióticas o distintos compuestos glicoconjugados, de interés creciente en la industrias alimentaria y farmacéutica.

Several epidemiological studies have shown that food plays a key role in human health. It has been shown that diet has a direct effect on the composition and balance of the intestinal microbiota and this, in turn, on health. Therefore, new "functional" foods are sought, many of which are prebiotic oligosaccharides that selectively stimulate the growth of beneficial bacteria.

On the other hand, the glycosylation of different chemical compounds improves their properties and their bioavailability. These new bioactive compounds have yet to be fully exploited, being increasingly demanded by the pharmaceutical industry.

Research on biosynthesis of these products involves the study of new compounds with better functional properties and, also, the development of more efficient production methods. Thus, Enzymatic Biotechnology has overcome other traditional chemical methodologies, due to its excellent regio-, chemo- and stereo-specificity. Furthermore, X-ray crystallography generates the basic structural information for rational design and protein engineering.

Accordingly, this thesis has focused on the crystallographic study of glycosyl hydrolases (GH), enzymes that metabolize carbohydrates, the most structurally diverse biocompounds in nature. These highly specific and efficient enzymes are involved in many essential biological processes. A detailed knowledge of their molecular function is crucial to understand these processes and control numerous diseases, and is necessary to optimize their biotechnological application. In this manuscript, we present the research work performed on four enzymes that can generate prebiotics of the types FOS (fructooligosaccharides), IMOS (isomaltooligosaccharides) or β -glucans. In addition, we have explored their ability to synthesize other bioactive compounds.

The general goal of the work is to shed light into the intricate mechanisms of protein-carbohydrate recognition that modulate the specificity of glycosidases, through protein crystallography. In addition to the fundamental implications for understanding the processes in which these enzymes are involved, this structural knowledge is crucial for performing the molecular engineering of enzymes. The ultimate goal is to produce oligosaccharides with prebiotic properties, specifically designed to manipulate the individual's intestinal microbiota, opening the door to personalized preventive medicine.

The first enzymes studied are the invertase from *Saccharomyces cerevisiae* (ScInv) and the β -fructofuranosidase from *Xanthophyllomyces dendrorhous* (XdInv), both belonging to the GH32 family. These enzymes are bimodular and are formed by a catalytic β -propeller domain and an adjacent β -sandwich domain. Our crystallographic studies have revealed that the β -sandwich domain is responsible for oligomerization and that this molecular assembly regulates the substrate specificity of yeast GH32 enzymes. While ScInv is an octamer that restricts access to long substrates, explaining its essentially invertase activity, XdInv presents a different dimer organization than previously described in the family. This distinct dimeric assembly could allow the degradation of branched mixed fructans, which might be related to its biological function in the natural habitat of the yeast.

ScInv and XdInv are able to generate oligosaccharides of type β (2-6)-FOS, with higher prebiotic capacities than the commercialized β (2-1)-FOS. Thus, ScInv produces 6-kestose and XdInv synthesizes neokestose. A detailed study of the XdInv active site, by analysis of complexes with various substrates, showed the molecular basis of its broad specificity and revealed the existence of two possible alternative binding sites, depending on the type of substrate. Based on these results, we explored the possibility of using other oligosaccharides as acceptor substrates. Thus, we found that the enzyme is capable of fructosylating maltooligosaccharides and produces hetero-FOS with

potential prebiotic capacities. Likewise, we have investigated the possibility of fructosylating other non-sugars organic compounds as antioxidants, to generate new bioactive compounds of interest.

Xanthophyllomyces dendrorhous α -glucosidase (Xd- α Glu) is an enzyme producing medium-length α (1-6)-IMOS, which posses improved functional and prebiotic properties. The structural model, generated by sequence homology, reveals that it presents a typical topology of eukaryotic GH31 enzymes. However, it has two novel structural features: an insert in the N-terminal domain, and an unusually long Insert 2 in the catalytic domain. These two new properties may be responsible for its preference for long substrates.

Finally, we studied a β -glycosidase isolated from the bovine rumen metagenome (GlyA₁). The GlyA₁ structure exhibits a domain permutation, but retains the general architecture of the GH3 β -glucosidases. We have shown that this new topology is associated with microorganisms that inhabit the digestive tract and that depend on a plant biopolymer rich diet. In addition, GlyA₁ has an extra C-terminal domain not previously described. Small-angle X-ray scattering (SAXS) data shows a compact overall structure where the C-terminal extra domain is positioned over the structural core of the enzyme.

GlyA₁ is a broad specificity β -glucosidase/ β -xylosidase, which also has other associated activities. This enzyme can degrade oligosaccharides with different type of binding and polymerization degree, being more active with medium-sized substrates. These characteristics make GlyA₁ a good candidate for several biotechnological applications. Specifically, its retention mechanism and its broad specificity suggest that it may be used to obtain different prebiotics and other biologically interesting oligosaccharides.

In conclusion, protein crystallography is a powerful tool to determine in detail the mechanism of action of enzymes. This thesis presents the structural study of three glycosyl hydrolase enzymes (ScInv, XdInv and GlyA₁) and their complexes, in addition to a computational model of another GH (Xd- α Glu). Our results have allowed us to delve into the sophisticated molecular mechanism that presents each enzyme to specifically recognize the substrate. In turn, this knowledge allows the rational design of enzymes with improved synthetic abilities, directed to the production of carbohydrates with prebiotic capacities or other novel glycoconjugate of increasing interest in the food and pharmaceutical industries.

ABREVIATURAS

ADN	Ácido desoxirribonucleico
Bis-Tris	Bis(2-hidroxietil)-amino-tris(hidroximetil)-metano
BTpro	Bis-Tris propano o 1,3-Bis[tris(hidroximetil)metilamino]propano
CAZy	Base de datos de enzimas activas en carbohidratos (<i>Carbohydrate Active Enzymes Database</i>)
CAZymes	Enzimas activas en carbohidratos (<i>Carbohydrate Active Enzymes</i>)
CBM	Módulo de unión a carbohidratos (<i>Carbohydrate Binding Module</i>)
Da	Dalton
DP	Grado de polimerización (<i>Degree of Polymerization</i>)
EC	Comisión de enzimas (<i>Enzyme Commission</i>)
EFSA	Autoridad europea en seguridad alimentaria (<i>European Food Safety Authority</i>)
FH	Fructosil hidrolasa
FOS	Fructooligosacáridos
FT	Fructosil transferasa
g	Gramo
GeOS	Gentiooligosacáridos
GH	Glicosil hidrolasa
Glu (Glup)	Glucosa (Glucopiranososa)
GlyA₁	β-Glicosidasa de metagenoma de rumen bovino
GOS	Galactooligosacáridos
GRAS	Ingredientes reconocidos como seguros (<i>Generally Recognized as Safe</i>)
GS	Glicosil sintasa
GuOS	Glucooligosacáridos
Fru	Fructosa
FDA	Administración de alimentos y medicamentos (<i>Food and Drug Administration</i>)
FnIII	Dominio fibronectina tipo III (<i>Fibronectin-like type III domain</i>)
FOSHU	Alimentos específicos para el uso en salud (<i>Foods of Specified Health Use</i>)
HEPES	Ácido 4-(2-hidroxietil)-1-piperazinaetanosulfónico
HGC	Grupo con alto recuento de genes (<i>High Gene Count</i>)

HMO	Oligosacáridos de la leche humana (<i>Human Milk Oligosaccharides</i>)
IMOS	Isomaltooligosacáridos
ISAPP	Asociación Científica Internacional para Probióticos y Prebióticos (<i>International Scientific Association for Probiotics and Prebiotics</i>)
IUB	Unión Internacional de Bioquímica (<i>International Union of Biochemistry</i>)
K	Grado Kelvin
kDa	Kilodalton
KOS	Kojioligosacáridos
L	Litro
LGC	Grupo con bajo recuento de genes (<i>Low Gene Count</i>)
LS	Lactosacarosa
M	Molar
mg	Miligramo
MFOS	Maltosilfructósidos
μm	Micrómetro
mM	Milimolar
MPD	2-Metil-2,4-pentanediol
m/v	Relación masa/volumen
NOS	Nigerooligosacáridos
ORF	Marco de lectura abierto (<i>Open Reading Frame</i>)
PAGE	Electroforesis en gel de poliacrilamida (<i>Polyacrylamide Gel Electrophoresis</i>)
PEG	Polietilenglicol
SAXS	Dispersión de rayos X a bajo ángulo (<i>Small angle X-ray scattering</i>)
ScInv	Invertasa de <i>Saccharomyces cerevisiae</i>
SCFA	Ácidos grasos de cadena corta (<i>Short Chain Fatty Acids</i>)
SoFfase	β-fructofuranosidasa de <i>Schwanniomyces occidentalis</i>
Tris	Tris(hidroximetil)aminometano
UA (ASU)	Unidad asimétrica (Asymmetric unit)
Xd-αGlu	α-Glucosidasa de <i>Xanthophyllomyces dendrorhous</i>
XdInv	β-Fructofuranosidasa de <i>Xanthophyllomyces dendrorhous</i>
XOS	Xilooligosacáridos

Resumen	I
Summary	III
Abreviaturas	V
Índice	1
1. Introducción	3
1.1) El papel de la alimentación en la salud	3
1.2) La microbiota intestinal.....	4
1.3) Los alimentos funcionales o nutraceuticos	6
1.3.1) La fibra dietética.....	6
1.3.2) Los prebióticos.....	7
1.4) Las glicosil hidrolasas.....	13
1.4.1) Clasificación de las glicosil hidrolasas	13
1.4.2) Mecanismo de acción de las glicosil hidrolasas.....	14
1.4.3) Estructura de las GH32	16
1.4.4) Estructura de las GH31	18
1.4.5) Estructura de las GH3.....	21
1.5) La Biotecnología Enzimática	24
Aportación de la Biología Estructural a la Biotecnología Enzimática.....	24
2. Objetivos	27
3. Enzimas productoras de FOS: Invertasa de <i>S. cerevisiae</i> y β-Fructofuranosidasa de <i>X. dendrorhous</i>.....	29
3.1) Introducción	29
3.2) Trabajo de investigación #1	31
<i>Three-Dimensional Structure of Saccharomyces Invertase. Role of a Non-Catalytic Domain in Oligomerization and Substrate Specificity.</i>	
3.3) Trabajo de investigación #2	45
<i>Structural Analysis of β-Fructofuranosidase from Xanthophyllomyces dendrorhous Reveals Unique Features and the Crucial Role of N-Glycosylation in Oligomerization and Activity.</i>	
3.4) Material inédito.....	62
3.5) Discusión.....	65
4. Enzimas productoras de IMOS: α-Glucosidasa de <i>X. dendrorhous</i>.....	69
4.1) Introducción	69
4.2) Trabajo de investigación #3	71
<i>Molecular Characterization and Heterologous Expression of a Xanthophyllomyces dendrorhous α-Glucosidase with Potential for Prebiotics Production.</i>	
4.3) Discusión.....	84
5. Enzimas productoras de β-glucanos, XOS y GeOS: β-Glucosidasa de Rumen	87
5.1) Introducción	87
5.2) Trabajo de investigación #4	89
<i>Structural and Functional Characterization of a Ruminant β-Glycosidase Defines a Novel Subfamily of Glycoside Hydrolase Family 3 with Permuted Domain Topology.</i>	
5.3) Discusión.....	106

6. Consideraciones finales.....	109
7. Conclusiones.....	111
8. Bibliografía	113
9. Anexos.....	123
Anexo I: Resumen oligosacáridos referidos en la tesis	123
Anexo II: Cristalización y resolución de proteínas.....	125
<i>Cristalización.....</i>	<i>126</i>
<i>Preparación de las muestras y recogida de datos de difracción.....</i>	<i>128</i>
<i>Concepto de cristal, celdilla unidad y unidad asimétrica.</i>	<i>129</i>
<i>Resolución estructural de la proteína</i>	<i>130</i>
<i>Cristalización de complejos proteína-ligando</i>	<i>131</i>
Anexo III: Otras técnicas de biología estructural.....	132
<i>Difracción de rayos X a bajo ángulo (SAXS)</i>	<i>132</i>
<i>Métodos de predicción de estructura terciaria</i>	<i>136</i>

1. INTRODUCCIÓN

1.1) EL PAPEL DE LA ALIMENTACIÓN EN LA SALUD

Las culturas occidentales viven en una paradoja: a pesar de la sobrealimentación de sus individuos, la dieta que adoptan no les aporta los nutrientes necesarios para la homeostasis y necesitan del aporte extra de ciertos alimentos para el desarrollo equilibrado del organismo.

Antes del desarrollo de la agricultura y la ganadería, la dieta de los homínidos estaba limitada a las plantas salvajes y a los animales que cazaban. Una vez que consiguieron domesticar a los animales y sembrar los cultivos, desarrollaron el procesamiento de los alimentos, como el queso, que aseguraban la presencia de nutrientes en la dieta a largo plazo. Este procedimiento se hizo más habitual a partir de la revolución industrial, donde los alimentos comenzaron a ser procesados y sanitizados de forma exhaustiva, hasta derivar en el consumo vigente¹: productos de manufactura rápida y barata, pero con valores nutricionales bajos. Este tipo de alimentación moderna ha alterado la cantidad y la clase de carbohidratos, lípidos y proteínas que se ingieren diariamente, y con ello, el aporte de nutrientes esenciales y beneficiosos para la salud del individuo.

Concretamente, la dieta occidental actual se caracteriza por un elevado consumo de carnes rojas, azúcares refinados y grasas saturadas, y una baja presencia de fibra, necesaria para el normal tránsito gastrointestinal. Este tipo de alimentación está relacionada con la obesidad, las enfermedades cardiovasculares, la hipertensión, la diabetes y el riesgo de padecer cáncer de colon. Estas enfermedades son una epidemia en las sociedades industrializadas actuales, afectando al 50-65% de la población, mientras que son poco habituales o inexistentes en los colectivos no occidentalizados, como los habitantes de zonas rurales y agrícolas². De ahí que a estas patologías se las conozcan como “enfermedades occidentales”¹. Como los cambios en el genotipo humano no pueden explicar el rápido incremento de estas patologías, la pérdida de los factores de protección y/o la aparición de nuevos factores de riesgo deben contribuir al aumento de la prevalencia de estas enfermedades³.

Desde hace varios años, se ha puesto la atención en las bacterias que colonizan nuestro intestino. La microbiota intestinal es, cuantitativamente, la fuente más importante de microbios en el cuerpo humano y su presencia celular es mayor al conjunto de células humanas. Por tanto, los cambios que se producen sobre ella deben tener repercusiones en el equilibrio y salud de los individuos. Así, se ha propuesto la teoría sobre “la afluencia de síndromes por privación de microorganismos”⁴. La variedad y número de bacterias que pueblan nuestro intestino se ha reducido con el paso de los siglos, y esta disminución se ve reflejada en una menor funcionalidad y protección del hospedador. Este empobrecimiento es debido a factores como el uso de antibióticos o la dieta³. Varios autores apoyan la idea de que incluso aquellos individuos que se consideran “sanos y saludables”, en realidad presentan una microbiota disfuncional y que podría predisponerlos a enfermedades. Por ello, encuentran necesario reconducir la microbiota a estados anteriores a la revolución industrial⁵, para lo cual es necesario el tratamiento mediante una dieta rica en fibra dietética y prebióticos, de los que se hablará en el apartado 1.3.

1.2) LA MICROBIOTA INTESTINAL

El término microbiota hace referencia a “la comunidad de microorganismos residentes en un nicho ecológico determinado”⁶. Además de contribuir a la digestión de los alimentos, principalmente carbohidratos, la microbiota intestinal complementa la nutrición del individuo mediante la producción de vitaminas⁷. También juega un papel importante en la regulación del metabolismo del hospedador, contribuyendo al desarrollo del sistema inmune así como en el desarrollo y maduración de la mucosa intestinal. Asimismo, la microbiota protege frente a la invasión de microorganismos patógenos⁸.

La población bacteriana en el tracto gastrointestinal difiere de sección en sección, debido a la variación de las condiciones ambientales a lo largo del tracto digestivo. En los individuos sanos, el estómago y el duodeno están poco colonizados (10-1000 células/mL), debido a la acidez extrema del entorno. La población bacteriana aumenta hacia la parte más distal del intestino delgado, consolidándose en el colon, donde alcanza una concentración máxima de un billón de células/mL⁹. La capa de mucosa fina y la rápida circulación de materia en el intestino delgado, sólo permite la degradación de los glicanos solubles¹⁰. Mientras, en el intestino grueso, el ecosistema es más complejo y numeroso gracias al pH cercano a la neutralidad, a la alta disponibilidad de nutrientes, así como al tránsito lento de los mismos¹¹. En este entorno, los glicanos disponibles son los menos solubles, y la microbiota intestinal coloniza la fibra dietética y la mucosa intestinal¹⁰.

Por otra parte, el intestino delgado presenta una fluctuación diaria en la composición de su ecología microbiana, reflejo del aporte cotidiano de distintos componentes³. Sin embargo, la composición de la microbiota del intestino grueso parece mantenerse constante a lo largo de los años. Es una coalición de dos filos mayoritarios, que se distribuyen en una continuidad. En un extremo, se encuentran predominantemente los miembros Gram-negativos del filo Bacteroidetes, que contienen los géneros *Bacteroides* y *Prevotella*, y en el otro extremo, se localizan preferentemente los miembros del filo Firmicutes, incluyendo el género *Lactobacillus* spp. El siguiente filo más abundante son las Actinobacterias, principalmente compuesto por el género *Bifidobacterium* spp¹². También están presentes, aunque en menor proporción, los filos Proteobacteria (que incluye *Escherichia coli*) y Verrucomicrobia (que incluye *Akkermansia muciniphila*, productor de butirato)¹³.

Desde los estudios de Tissier H. en 1900, se acepta que el feto es estéril *in utero*, y que los bebés nacen sin microbiota. Sin embargo, este sistema ecológico tan importante para los mamíferos se adquiere rápidamente tras el parto⁸. La microbiota materna forma el primer inóculo y, tras el nacimiento, la diversidad bacteriana aumenta hasta llegar a una microbiota “adulta”, en términos de composición y diversidad, que converge a los 3-5 años de vida³. Una vez establecida la “madurez” de la microbiota, su composición se mantiene relativamente estable a lo largo de la vida del individuo, si bien puede resultar alterada por procesos infecciosos, el uso de antibióticos, el estilo de vida, algunas cirugías y el tipo de dieta a largo plazo².

Recientemente, se ha descubierto que un tercio de la microbiota es heredada, mientras que los otros dos tercios son consecuencia de los hábitos alimentarios, la higiene, el tratamiento con antibióticos y la genética humana¹⁴. Por ello, estos primeros años son un periodo crítico a nivel nutricional, y alteraciones durante esta etapa pueden derivar en profundos déficits para la salud y el desarrollo del ser humano, así como en la aparición de distintas enfermedades en los siguientes años de vida.

La microbiota intestinal ejerce una gran interacción con el individuo, a través de su epitelio intestinal. La mucosa intestinal actúa de barrera física contra las bacterias invasoras, pero también sirve de anclaje a las bacterias que viven en simbiosis en este hábitat. Aquellos microorganismos que son capaces de usar los glicósidos del mucus como fuente de carbono, se sitúan más cerca de las

células del hospedador. Como consecuencia, durante estados de disbiosis, estos microorganismos ejercen una degradación desproporcionada de los glicanos endógenos humanos, causando problemas gástricos¹⁰. La disbiosis (o disbacteriosis), es un inadecuado cambio de la microbiota bacteriana del organismo humano, tanto en composición como en actividad, debido a una enfermedad y/o alteración de las condiciones del hospedador², como puede ocurrir tras el uso de antibióticos, cambios en la dieta u otros factores. Si esta situación de disbiosis se mantiene en el tiempo, se incrementa el número de microorganismo patógenos, provocando un aumento de los procesos inflamatorios y la respuesta inmune, pudiendo llegar a invadir el epitelio intestinal. Además, tan importante es la influencia de la alimentación en la salud, y tan desajustada es la dieta moderna de las necesidades nutricionales reales del individuo que algunos autores han sugerido que la microbiota de los individuos occidentales en realidad es disbiótica y predispone al individuo a una serie de enfermedades relacionadas con la inflamación y la respuesta inmune, pudiendo contribuir a la obesidad y a otros desordenes metabólicos, como la diabetes tipo II^{2,5} (figura 1.1).

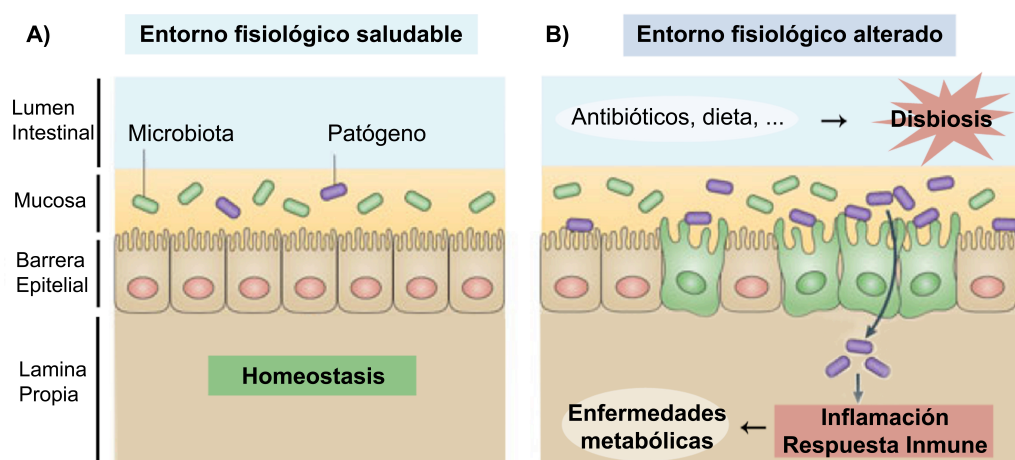


Figura 1.1. Comparación del papel de la microbiota intestinal en (A) un entorno fisiológico sano o en (B) condiciones alteradas que causan una disbiosis de la microbiota y una desprotección del epitelio intestinal. Modificada de ¹⁵.

En 2014, Knoghts D. y colaboradores¹⁶ confirmaron que los cambios ambientales, como un cambio en la dieta, afectan a la composición de la microbiota y, por ende, a sus funciones sobre el individuo. Paralelamente, Le Chatelier E. y colaboradores¹⁷ vieron que la riqueza genética del microbioma estaba fuertemente relacionada con los niveles de inflamación, con el metabolismo lipídico y con las respuestas a las variaciones de la dieta. Así, crearon dos conjuntos: el grupo con bajo recuento de genes (LGC, *Low Gene Count*) y el grupo con alto recuento de genes (HGC, *High Gene Count*). Aquellos individuos LGC presentaban mayores niveles de inflamación y peores respuestas en la adaptación de la dieta. En consecuencia, un simple test molecular basado en el genoma de nuestra microbiota intestinal podría servir de diagnóstico sobre el riesgo de desarrollar ciertas enfermedades. Es más, la riqueza genética de la microbiota intestinal del individuo puede servir para predecir el potencial de una intervención en su dieta, e incluso seleccionar la composición de la dieta más adecuada para cada persona⁴.

Por ello, resulta necesario un cambio radical que asuma la microbiota como un elemento esencial para la salud humana, empezando por realizar una modificación de los hábitos alimentarios de la población occidental. Sin embargo, esta medida puede no ser suficiente y actualmente se está poniendo la atención en el desarrollo de nuevos productos que ayuden a paliar este déficit microbiano.

1.3) LOS ALIMENTOS FUNCIONALES O NUTRACÉUTICOS

El término alimento funcional se utiliza para referirse a un alimento que tiene un demostrado efecto beneficioso sobre la salud, mejorando aspectos fisiológicos o reduciendo el riesgo de enfermedad crónica, más allá de la simple nutrición¹⁸. Si bien su uso es principalmente preventivo, son consumidos como parte habitual de la dieta. Por otro lado, los nutraceuticos son compuestos derivados de alimentos, que suelen ser comercializados como píldoras, cápsulas, líquidos o dosificados, y presentan efectos beneficiosos para la salud¹⁹.

En algunos países, ambos conceptos son intercambiables. Estos alimentos o ingredientes incluyen los ácidos grasos omega 3, los flavonoides o la fibra dietética¹⁹. Dentro de estos alimentos funcionales, la fibra dietética y los prebióticos son los que han despertado un mayor interés en los últimos años, probablemente debido a la gran incidencia de patologías relacionadas con la función gastrointestinal y las enfermedades metabólicas²⁰.

1.3.1) LA FIBRA DIETÉTICA

En 1953, Hipsley B. R. acuñó el término fibra dietética para referirse a los constituyentes no digeribles que componen la pared celular vegetal²¹. Todavía se está debatiendo la definición y especificación de la fibra dietética, pero de acuerdo con la Comisión CODEX Alimentarius, al igual que pasa con las vitaminas, la fibra alimentaria está compuesta por distintos tipos de moléculas que varían en su estructura y su función, así como en la cantidad mínima necesaria para un papel óptimo en la salud del individuo. En su última reunión en 2014, dentro de la fibra dietética se incluyen “aquellos polímeros de carbohidratos de tres o más unidades monoméricas, que no son digeridos ni absorbidos en el intestino delgado humano, pero que pueden ser fermentados en el intestino grueso”. Este término incluye lignina y componentes asociados con la fibra dietética, como los prebióticos²².

La fibra dietética está compuesta por una mezcla de polisacáridos que componen la pared celular vegetal y es un constituyente normal de la mayoría de los alimentos derivados de plantas²³. En la figura 1.2 se representa la estructura y composición de la pared celular primaria de las plantas.

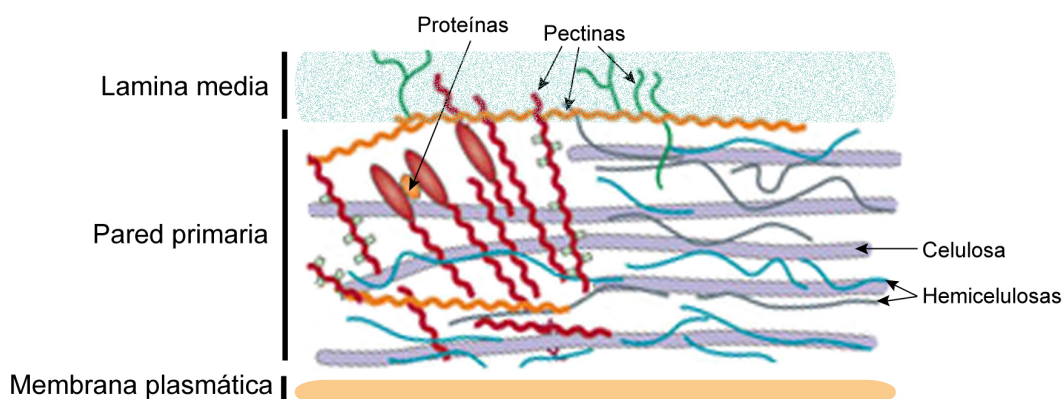


Figura 1.2. Composición de la pared primaria vegetal. Las microfibrillas de celulosa forman la matriz sobre la que se establecen otros polisacáridos como hemicelulosas y pectinas. En este entramado de carbohidratos también se encuentran presentes proteínas solubles. Todos estos elementos configuran y estructuran la robusta pared vegetal.

El componente mayoritario de las paredes vegetales es la celulosa, un polímero de glucosas unidas por enlace $\beta(1-4)$. La mayor parte de la celulosa es resistente a la fermentación por la microbiota intestinal debido a su estructura molecular (microfibrillas) y su asociación con otras sustancias como la pectina o la hemicelulosa, así como a algunas proteínas, lo que limita aún más su capacidad para ser digerida en el colon²¹.

La hemicelulosa se encuentra en la pared vegetal de plantas superiores. Este componente de la fibra dietética incluye xiloglucanos, arabinoxilanos, fucomananos, galactomananos y β -glucanos. La característica común de todos estos compuestos es el enlace $\beta(1-4)$ de su esqueleto, y la mayoría presenta ramificaciones (simples o complejas). Como excepción, los β -glucanos son polímeros lineales que también pueden tener enlaces $\beta(1-3)$ ²⁴.

La fibra dietética se puede dividir en fibra soluble y fibra insoluble. La parte soluble se disuelve en agua y, aunque elude la digestión en el intestino delgado, es fácilmente fermentada por la microbiota en el intestino grueso. Está formada por pectinas, gomas, fructanos y algunas hemicelulosas, como los β -glucanos. Por otra parte, la fibra insoluble no se disuelve en agua, su fermentación es más limitada, y se compone, fundamentalmente, de fragmentos de las paredes celulares que contienen celulosa, lignina y hemicelulosas más complejas²⁴. Así, mientras que una parte de la fibra sufre hidrólisis y fermentación en el colon, otra parte, también denominada fibra bruta, permanece inalterada y es de gran importancia como inductora del peristaltismo intestinal²¹. La mayoría de las fibras contienen aproximadamente un tercio de fibra soluble y dos tercios de fibra insoluble⁹.

En general, los procesos de extracción de fibra parten de subproductos de otras operaciones alimentarias. La fibra puede ser extraída en conjunto, dividida en soluble e insoluble, o separando sus constituyentes individuales, y dependiendo del tipo de fibra que se quiera obtener, el proceso de extracción es distinto. La técnica más tradicional implica una desintegración física de las distintas fuentes de fibra (molido y extrusión). También se puede usar la extracción química, que implica tratamientos con peróxido de hidrógeno o con carbonato de sodio, entre otros compuestos. Los tratamientos enzimáticos conducen a incrementos en el contenido de la fibra soluble. Además, este tipo de fibra tiene una textura más suave, lo cual facilita su uso en la formulación de productos²³. Por tanto, para obtener mejores características y obtener una fibra dietética más efectiva es necesario estudiar nuevas enzimas capaces de modificar los componentes de la pared celular de las plantas. El uso combinado de las β -glucosidasas y las β -xilosidasas podría producir una fibra dietética más soluble y con mejores prestaciones.

1.3.2) LOS PREBIÓTICOS

En general, la fibra dietética son carbohidratos que en el colon sirven como sustrato para los microorganismos que lo habitan, originando sustratos metabólicos y micronutrientes esenciales para el hospedador, como los ácidos grasos de cadena corta (SCFA, *Short Chain Fatty Acids*). Los prebióticos realizan todas las actividades mencionadas anteriormente, pero además, estimulan el crecimiento selectivo de determinadas especies beneficiosas de la microbiota intestinal²⁵.

Los primeros estudios sobre prebióticos comenzaron en los años 80, cuando investigadores japoneses demostraron en cultivos *in vitro* que ciertos oligosacáridos no digeribles eran fermentados selectivamente por bifidobacterias y que, además, estos compuestos tenían la capacidad de estimular su crecimiento²⁶. Estos resultados fueron confirmados por Gibson G. R. y Roberfroid M. en 1995²⁵ quienes, conjuntamente, definieron por primera vez el término prebiótico. Actualmente, la definición más utilizada es la concretada por la Asociación Científica Internacional para Probióticos y Prebióticos (ISAPP, *International Scientific Association for Probiotics and Prebiotics*). Así, se

reconocen como prebióticos a “aquellos ingredientes alimentarios que al ser fermentados selectivamente producen cambios específicos en la composición y/o actividad de la microbiota gastrointestinal confiriendo beneficios en la salud del individuo”²⁷.

No todos los carbohidratos de la dieta son prebióticos. Para que un compuesto sea considerado prebiótico debe: 1) ser resistente a la acidez gástrica, a la hidrólisis por las enzimas del hospedador y a la absorción gastrointestinal, 2) ser fermentado por la microbiota intestinal, 3) ser capaz de estimular selectivamente la composición y/o actividad de dicha microbiota intestinal, hacia un ecosistema más saludable e 4) inducir efectos fisiológicos beneficiosos para la salud del individuo²⁷.

Estos carbohidratos prebióticos pueden estar presentes de forma natural en alimentos tales como la leche y la miel, así como en hortalizas, verduras, frutas, cereales, legumbres y frutos secos, de donde se pueden extraer mecánicamente de forma semejante a la descrita en la fibra dietética. También pueden obtenerse por degradación de polisacáridos, utilizando métodos químicos o enzimáticos²⁸. Además, los prebióticos se pueden producir mediante síntesis química o enzimática. Mediante síntesis química pueden obtenerse oligosacáridos prebióticos como la lactulosa, un disacárido que se obtiene por isomerización de la lactosa utilizando catalizadores básicos¹¹. Para la síntesis enzimática se utilizan diferentes tipos de enzimas, dependiendo del prebiótico que se quiera producir, que pueden proceder de diferentes microorganismos (hongos, levaduras y bacterias). En la figura 1.3 se resumen los procesos de producción de prebióticos.

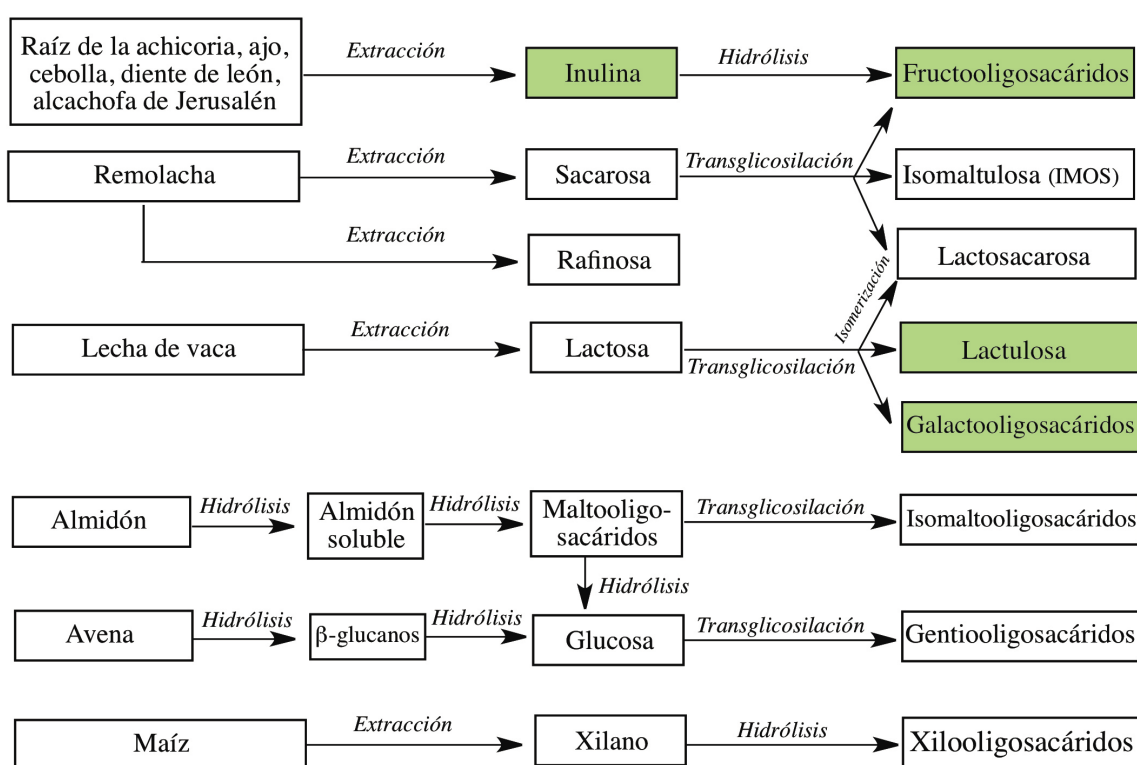


Figura 1.3. Recopilación de los procesos utilizados para producir oligosacáridos prebióticos (señalados en verde) y otros carbohidratos potencialmente prebióticos. Adaptado de ¹¹.

A día de hoy, sólo se reconocen como prebióticos los oligosacáridos de la leche humana (HMO, *Human Milk Oligosaccharides*), los galactooligosacáridos (GOS), la lactulosa y los fructanos (inulina y los fructooligosacáridos, FOS). Sin embargo, existe un gran interés por comercializar nuevos oligosacáridos prebióticos con propiedades funcionales mejoradas. Dentro de los denominados “prebióticos emergentes”, algunos ya están disponibles comercialmente, a pesar de no contar con todos los resultados necesarios para justificar su capacidad prebiótica. Entre ellos se encuentran los xilooligosacáridos (XOS), los gentiooligosacáridos (GeOS), la lactosacarosa (LS), los oligosacáridos de

la soja (rafinosa y estaquiosa) y los isomaltooligosacáridos (IMOS)¹¹. Otros carbohidratos, aún más nuevos, todavía están en una fase inicial de estudio. Cabe destacar los β -glucanos, ya que se ha visto que son capaces de estimular el crecimiento de bifidobacterias en el colon humano²¹.

En la actualidad, los prebióticos se añaden como ingredientes a una gran variedad de productos distintos como bebidas, lácteos, galletas, sopas, cereales y patés, además de usarse como suplementos dietéticos. En 2015, el mercado mundial de prebióticos superó los 3300 millones de dólares, y se calcula que este consumo se duplicará en los próximos 10 años²⁹.

A continuación se detallarán las características de los oligosacáridos implicados en este estudio. En el Anexo I se encuentra una relación de todos los carbohidratos mencionados en este manuscrito, así como su nombre sistemático.

A) INULINA Y FRUCTOOLIGOSACÁRIDOS (FOS)

Los fructanos son polímeros de unidades de fructosa. Si el polisacárido está formado por fructosas unidas por enlaces $\beta(2-1)$ se denomina inulina, mientras que si están unidas por enlaces $\beta(2-6)$ se conoce como levano. Por último, existen polisacáridos mixtos, llamados graminanos, que contienen enlaces $\beta(2-1)$ y $\beta(2-6)$ ³⁰.

Los fructooligosacáridos (FOS) son oligosacáridos de cadena corta (de hasta 9 unidades) muy solubles y con propiedades tecnológicas beneficiosas, como la alta viscosidad, que proporciona una mejor textura y sensación en la boca. Además, son compuestos bajos en calorías y no cariogénicos³¹. Asimismo, presentan una alta eficiencia en la inhibición del crecimiento de ciertos microorganismos que contaminan la comida durante su almacenamiento, equivalente a la que presenta el sorbitol³². Por ello, a pesar de su menor dulzor, se están empezando a usar como sustituyente de la sacarosa en la industria alimentaria³³.

Estos oligosacáridos se pueden obtener por hidrólisis de los fructanos o por transfructosilación enzimática. Mediante el uso de endo-inulinasas y endo-levanasas se produce una hidrólisis parcial del polímero que da lugar a una mezcla de oligofructanos que pueden presentar grados de polimerización de 2 a 7 unidades, con un promedio de 4. Sin embargo, la síntesis enzimática de FOS por transfructosilación, utilizando sacarosa como sustrato, da lugar a productos más homogéneos¹¹.

Dependiendo del tipo de enlace y del aceptor de la fructosa, los FOS se pueden clasificar en cuatro tipos (figura 1.4):

- FOS tipo inulina o ¹F-FOS: presentan enlaces $\beta(2-1)$ entre las fructosas, con una unidad de glucosa al final. A este grupo pertenece la 1-kestosa, la nistosa y la fructosil-nistosa³¹.
- FOS tipo levano o ⁶F-FOS: presentan enlaces $\beta(2-6)$ entre las fructosas, con una unidad de glucosa al final. A este grupo pertenece la 6-kestosa³⁴.
- Neo-fructooligosacáridos o neo-FOS (¹G-GOS y ⁶G-FOS): se producen por la fructosilación de la unidad de glucosa de la sacarosa o la 1-kestosa, por enlaces $\beta(2-1)$ o $\beta(2-6)$. Este es el caso de la neokestosa y la neonistosa³².
- Hetero-fructooligosacáridos o hetero-FOS: este grupo incluye aquellos fructooligosacáridos formados por la fructosilación de azúcares distintos de la sacarosa. La fructosilación de la maltosa produce maltosilfructósidos (MFOS) como la erlosa y la neoerlosa. Dentro de este apartado también se incluyen la lactosacarosa (LS)³⁵.

Se ha visto que los FOS con enlaces $\beta(2-6)$ presentan mayores beneficios para la microbiota intestinal, y por tanto, para la salud del individuo. Así, en comparación con los FOS tipo inulina, los FOS tipo levano promueven un mayor crecimiento de *Bifidobacterium lactis* e inducen una mayor

síntesis de SCFA³⁶. Además, los neo-FOS también presentan un mayor efecto prebiótico que los FOS comercializados $\beta(2-1)$ ³⁷.

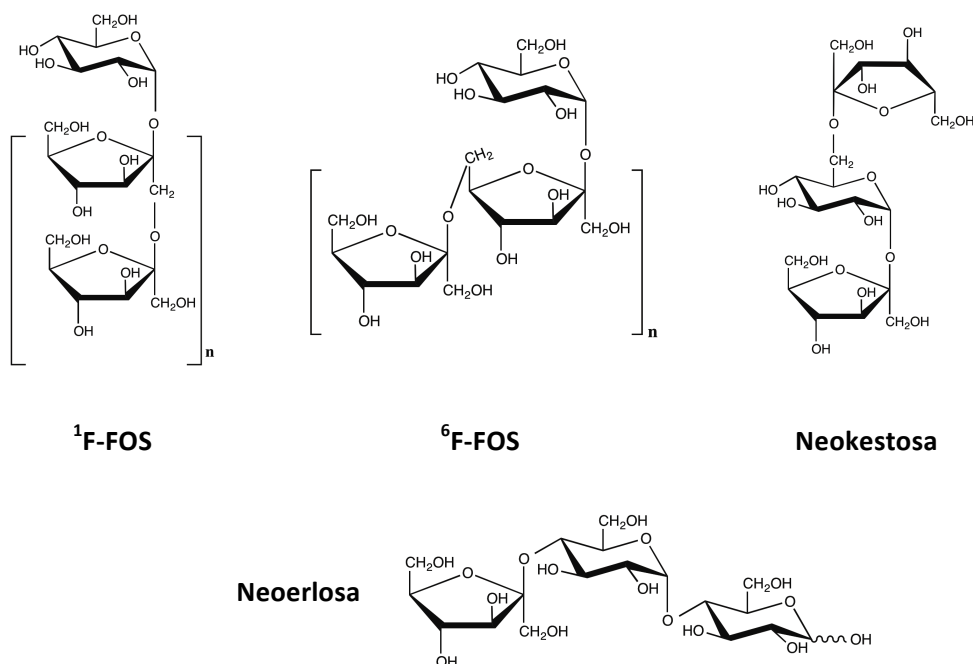


Figura 1.4. Esquema de las estructuras de los distintos tipos de FOS.

Está aceptado que la inulina y los FOS (¹F-FOS, ⁶F-FOS y neo-FOS) no se degradan ni se absorben en el tracto gastrointestinal superior, de tal forma que llegan intactos al colon donde son metabolizados por la microbiota intestinal. La configuración beta del carbono anomérico de la fructosa les hace resistentes a la hidrólisis por las enzimas digestivas humanas³⁸⁻³⁹. Además, diferentes ensayos de fermentación *in vitro*, así como ensayos en humanos, han puesto de manifiesto que la inulina y los FOS favorecen el crecimiento de bifidobacterias y lactobacilos⁴⁰. Esa especificidad de las bifidobacterias por los FOS debe producirse por la alta producción de β -fructofuranosidasas de estas especies²⁵. Además, se ha demostrado que la inulina y los FOS reducen el crecimiento de potenciales patógenos como *E. coli*, *Salmonella* y *Listeria*^{24, 37}. Por otro lado, los MFOS y el resto de hetero-FOS aún tienen que ser examinados para asegurar su capacidad prebiótica.

Cabe destacar que la asignación de la inulina y los FOS como ingredientes reconocidos como seguros o alimentos GRAS (*Generally Recognized as Safe*) por la FDA (*Food and Drug Administration*) en EEUU y como FOSHU (*Foods of Specified Health Use*) en Japón, ha permitido que actualmente se utilicen sin restricciones en un gran número de alimentos¹¹.

B) ISOMALTOOLIGOSACÁRIDOS (IMOS)

Los isomaltooligosacáridos (IMOS) son carbohidratos de cadena corta (de 2 a 9 unidades) con distintos tipos de enlaces α entre glucosas. En un sentido estricto, los IMOS son glucosacáridos con enlaces $\alpha(1-6)$, como la isomaltosa. Sin embargo, actualmente también se aceptan aquellos glucooligosacáridos que contienen enlaces $\alpha(1-6)$ y $\alpha(1-4)$, como la panosa. No obstante, en el término IMOS se extiende a gran cantidad de glucosacáridos distintos. A continuación se detallan los

distintos tipos de isomaltooligosacáridos y la denominación que se utilizará para el resto del manuscrito⁴¹ (figura 1.5):

- Isomaltooligosacáridos (IMOS): glucooligosacáridos lineales con enlaces $\alpha(1-6)$ y $\alpha(1-4)$.
- Nigeroooligosacáridos (NOS): glucooligosacáridos lineales con enlaces $\alpha(1-3)$, y algunos también presentan enlaces $\alpha(1-6)$.
- Kojioligosacáridos (KOS): glucooligosacáridos lineales con enlaces $\alpha(1-2)$, y algunos presentan también enlaces $\alpha(1-6)$.
- Glucooligosacáridos (GuOS) y oligodextranos: los GuOS son oligosacáridos ramificados con enlaces $\alpha(1-6)$ y $\alpha(1-4)$, producidos por las dextranacarasas a partir de sacarosa. Mientras, los oligodextranos son glucooligosacáridos lineales o ramificados producidos por hidrólisis controlada de los dextranos. Principalmente contienen enlaces $\alpha(1-6)$ y $\alpha(1-4)$, pero también pueden contener enlaces $\alpha(1-2)$ o $\alpha(1-3)$.
- Hetero-isomaltooligosacáridos o hetero-IMOS. Son glucosas unidas a otros monómeros o compuestos mediante un enlace $\alpha(1-6)$. La isomaltulosa, comercialmente conocida como Palatinosa™, es un disacárido formado por una glucosa unida a una fructosa mediante un enlace $\alpha(1-6)$. Los IMOS-alditoles son hetero-IMOS producidos por dextranacarasas que glucosilan alditoles como manitol y sorbitol.

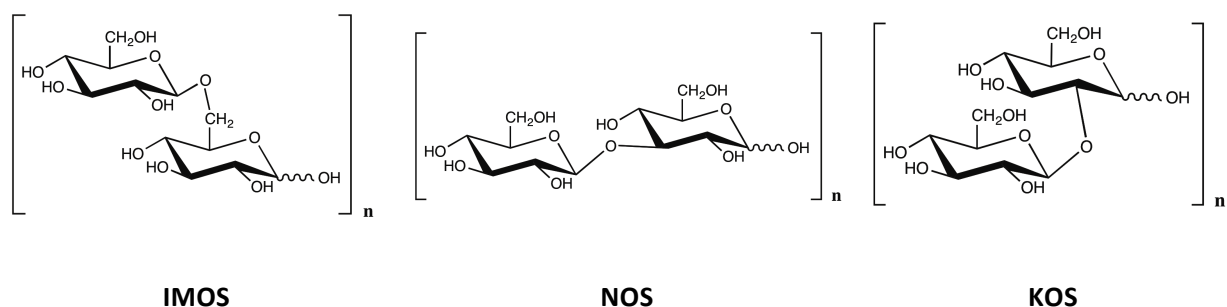


Figura 1.5. Esquema de las estructuras de distintos tipos de isomaltooligosacáridos.

Los IMOS son oligosacáridos funcionales bien conocidos en Asia, y emergentes prebióticos en Europa y América. Se encuentran de forma natural en varios alimentos fermentados como el miso, el sake y la salsa de soja, aunque también en alimentos no procesados como la miel. Sin embargo, su extracción no resulta económica⁴¹, por ello es necesario buscar nuevos procesos enzimáticos para producir IMOS de forma eficiente. A día de hoy, existen varias estrategias. La principal fuente de producción de IMOS es el almidón, un polisacárido de almacenamiento en las plantas¹⁰. A nivel industrial, se generan mediante la acción de α -transglucosidasas o glicosil transferasas (EC 2.4.1.24) de *Aspergillus* sp. sobre hidrolizados de almidón (maltosa y maltodextrinas). Debido a la naturaleza del sustrato y al proceso de transglucosilación, estos siropes de IMOS presentan gran variedad de compuestos con distinto grado de polimerización, algunos de ellos sin capacidad prebiótica y que deben ser eliminados⁴¹.

Por otro lado, se pueden usar las glicosil hidrolasas para sintetizar IMOS a partir de almidón, en un proceso que consta de dos etapas. En la primera etapa, las enzimas α -amilasa (EC 3.2.1.3) y pululanasa (EC 2.4.1.5) generan $\alpha(1-4)$ -maltooligosacáridos. Posteriormente, las α -glucosidasas (EC 3.2.1.20) realizan una transglucosilación, generando mayoritariamente $\alpha(1-6)$ -isomaltooligosacáridos, aunque dependiendo de la α -glucosidasa utilizada también puede presentar enlaces $\alpha(1-2)$ o $\alpha(1-3)$. Esta síntesis genera productos con menos compuestos secundarios y, por tanto, más fáciles de purificar²⁵. Por otra parte, los hetero-IMOS pueden generarse por transglucosilación, mediante el uso de las α -transglucosidasas o de las α -glucosidasas. Su producción está actualmente en estudio y desarrollo, así como sus capacidades prebióticas efectivas.

Dentro de las características de los IMOS como prebióticos, es importante estudiar su digestibilidad y su capacidad prebiótica. Se ha visto que los IMOS sufren hidrólisis parcial por las isomaltasas en el yeyuno humano (parte del intestino delgado), aunque quedan oligosacáridos que son fermentados por la microbiota intestinal²⁵. Se ha comprobado que IMOS con mayor grado de polimerización son menos digeribles y que los oligómeros ramificados son resistentes a las enzimas gastrointestinales⁴¹.

Los IMOS, los KOS, los NOS, los GuOs y la isomaltulosa han demostrado su capacidad bifidogénica en varios estudios, tanto *in vitro* como *in vivo*, en animales y/o en humanos. No obstante, la capacidad prebiótica del resto de compuestos (oligodextrano y otros hetero-IMOS) aún no ha sido probada⁴¹⁻⁴². Además, se ha visto que existe relación entre el tipo de enlace y la capacidad prebiótica: $\alpha(1-2) > \alpha(1-6) > \alpha(1-3) > \alpha(1-4)$ ⁴².

Además de la capacidad bifidogénica, los IMOS tienen otras características beneficiosas para la salud, como mejorar el tránsito intestinal (se ha demostrado que son más eficientes que los FOS para prevenir el estreñimiento), ayudan a la absorción de calcio, disminuyen el índice glucémico, y reducen los niveles de triglicéridos y ácidos grasos en sangre. Asimismo, los IMOS tienen un efecto inhibidor de las α -glucosidasas humanas y, por ello, se han usado como tratamiento terapéutico para disminuir la digestión del almidón y la sacarosa en enfermedades como la obesidad, la diabetes mellitus, la gastritis y las úlceras intestinales, y las caries. Por último, los IMOS presentan otras propiedades técnicamente funcionales, como prevenir el deterioro por microbios y retardar la degradación de los alimentos, además de ser dulces pero bajos en calorías^{11, 41}.

C) β -GLUCANOS

Los β -glucanos contienen un grupo heterogéneo de polisacáridos que, dependiendo del origen, presentan distintos enlaces y algunos muestran ramificaciones. Los β -glucanos de plantas son polímeros lineales que tienen enlaces $\beta(1-3)$ y enlaces $\beta(1-4)$, cuya distribución depende de la especie. La estructura común del polímero está constituida por bloques de dos o tres residuos adyacentes de β -glucosas unidos por enlaces $\beta(1-4)$, separados por un residuo de glucosa unido por enlace $\beta(1-3)$ ⁴³. En la figura 1.6 se muestra un ejemplo de β -glucano, el tetraglucano tipo B.

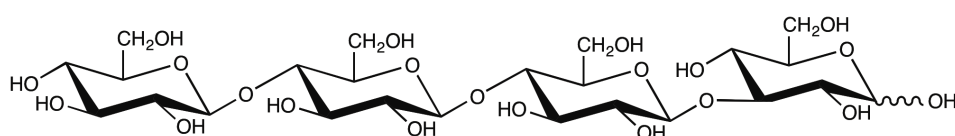


Figura 1.6. Esquema de la estructura del tetraglucano B (Glu- $\beta(1-4)$ -Glu- $\beta(1-4)$ -Glu- $\beta(1-3)$ -Glu).

En los últimos años ha aumentado el interés por los β -glucanos como posibles prebióticos, ya que estudios *in vitro* demuestran que son capaces de estimular el crecimiento de bifidobacterias y lactobacilos. Además, su configuración β lo hace indigerible por las enzimas humanas²¹. Por otra parte, se ha visto que son capaces de estimular la producción de SCFA, sobre todo la producción de butirato, un compuesto antiinflamatorio⁴⁴. También se ha visto que los β -glucanos con menor masa molecular producen más SCFA⁴³, mientras que los β -glucanos con mayores masas (230 kDa, más de 1000 glucosas) no producían variación en la microbiota²¹. Por tanto, la degradación parcial de los β -glucanos mediante β -glucosidasas aumentaría su capacidad funcional sobre la microbiota intestinal.

A nivel comercial, en 1997 la FDA aprobó el uso de β -glucanos para reducir el colesterol y disminuir el riesgo de padecer enfermedades coronarias. Recientemente, la EFSA (*European Food Safety Authority*) también ha reconocido que el consumo regular de β -glucanos de avena disminuye el colesterol en sangre⁴⁵.

1.4) LAS GLICOSIL HIDROLASAS

Las investigaciones sobre métodos de producción de estos oligosacáridos conllevan tanto el estudio de nuevos compuestos con novedosas propiedades funcionales como la búsqueda de formas más eficaces de producirlos. En este aspecto, la biotecnología enzimática se ha impuesto frente a otras metodologías de química tradicional (basadas en numerosos pasos de protección/desprotección de grupos funcionales), gracias a su excelente regio-, quimio- y estéreo-especificidad.

Las enzimas que son capaces de degradar, modificar o crear enlaces glicosídicos se denominan CAZymes (Enzimas Activas en Carbohidratos, *Carbohydrate Active Enzymes*)⁴⁶. Dentro de este grupo, se encuentran las glicosidasas o glicosil hidrolasas (GH), enzimas capaces de hidrolizar estos enlaces. Las mismas enzimas, en condiciones de alta concentración de sustrato, son capaces de llevar a cabo la reacción sintética contraria, la transglicosilación. Incluyen una gran variedad de patrones estructurales diferentes, así como distintas actividades específicas, y se encuentran en todos los reinos⁴⁷.

1.4.1) CLASIFICACIÓN DE LAS GLICOSIL HIDROLASAS

Atendiendo a la reacción que catalizan, según el Catálogo de Enzimas (EC, *Enzyme Catalog*) de la Unión Internacional de Bioquímica (IUB, *International Union of Biochemistry*), las glicosil hidrolasas se clasifican como EC 3.2.1.x, donde los tres primeros dígitos indican la propiedad de hidrolizar el enlace O-glicosídico, y el cuarto (x) indica el tipo de sustrato. Sin embargo, esta clasificación no refleja las características estructurales de cada enzima, por lo que no aporta información sobre las relaciones filogenéticas, perdiendo así información sobre eventos de evolución como la divergencia (enzimas con la misma estructura pero distinta especificidad de sustrato) o de convergencia (enzimas con distinta arquitectura pero que degradan el mismo tipo de sustrato)⁴⁸.

Por ello, a finales de los años 80, Henrissat B.⁴⁸ estableció una clasificación de las glicosidasas basada en la similitud de sus secuencias de aminoácidos. En aquel momento, se catalogaron 291 GH en 35 familias. Actualmente existen más de 300.000 secuencias, ordenadas en 135 familias. Además, el análisis comparativo de las estructuras terciarias de distintas familias reveló que existen familias que comparten la misma arquitectura catalítica, por lo que se clasifican dentro del mismo clan. En la actualidad, existen 14 clanes GH (designados de la A hasta la N). Una versión actualizada de esta clasificación está disponible *on-line* en la base de datos CAZy (*Carbohydrate-Active Enzymes Database*, www.cazy.org).

En algunos casos, este sistema de clasificación permite predecir el sustrato, en base a los carbohidratos asociados a dicha familia. Sin embargo, cuando una familia contiene muchas especificidades distintas, predecir la funcionalidad de una nueva enzima es muy complicado. Por ello, cuando es factible, dentro de cada familia aparecen subfamilias asociadas con cada tipo de actividad⁴⁶.

Los miembros de una misma familia mantienen una arquitectura mínima de dominios y comparten un mismo mecanismo catalítico. Sin embargo, muchas GH son grandes enzimas multidominio y presentan dominios adicionales, algunos con funciones todavía desconocidas, que los convierten en objetivos muy interesantes para estudios estructurales⁴⁹.

Como dato curioso, es interesante destacar que el genoma humano codifica para 97 glicosil hidrolasas, de las cuales, 17 están relacionadas directamente con la degradación de carbohidratos de la dieta (GH1, GH9, GH13, GH18, GH31, GH35 y GH37). Los 80 genes restantes codifican para enzimas

que no están asociadas a la digestión de alimentos, como por ejemplo, la síntesis de N-glicosilaciones⁵⁰. Este es un número pequeño en comparación con algunas de las bacterias intestinales, como *Bacteroides thetaiotaomicron*, cuyo genoma codifica más de 260 GH⁵¹. Por lo tanto, con la excepción del almidón, la lactosa y la sacarosa, la capacidad humana para digerir los carbohidratos de la dieta reside indiscutiblemente en su microbiota intestinal⁵².

Dentro de las GH expresadas por la microbiota e implicadas en la degradación de glicanos, el 40% están implicadas en la degradación de la pared celular de plantas, el 20% en la degradación de fructanos y el 10% en la degradación de almidón y glucógeno. Cabe destacar que GH13, implicada en la degradación de almidón, es la familia con más representantes (figura 1.7).

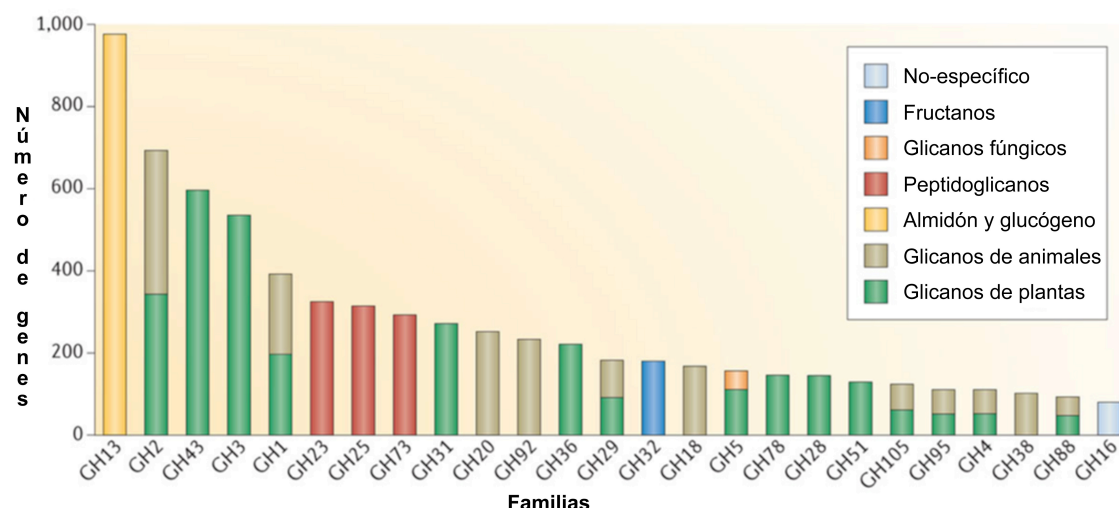


Figura 1.7. Abundancia de genes en las familias de glicosil hidrolasas (GH) expresada en el mini-microbioma, indicando por colores los sustratos sobre los que actúan. Adaptada de ⁵⁰.

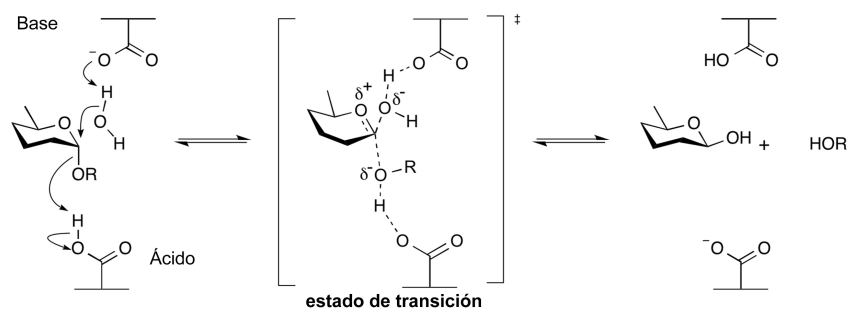
1.4.2) MECANISMO DE ACCIÓN DE LAS GLICOSIL HIDROLASAS

La hidrólisis del enlace glicosídico puede ocurrir de dos formas estereoquímicas posibles: la inversión o la retención de la configuración anomérica del sustrato. Es por ello que existen dos mecanismos de acción, que clasifican a estas enzimas en glicosidasas de inversión o de retención (figura 1.8). En el caso de las glicosidasas de inversión, dos residuos carboxílicos actúan como ácido y base, respectivamente, estando a unos 10,5 Å, para permitir la unión del sustrato y de una molécula de agua (figura 1.8.A). Esta reacción ocurre en un único paso, con un estado de transición de tipo ión oxocarbenio. Por otro lado, en las enzimas de retención, los grupos carboxílicos están a unos 5,5 Å de distancia y su acción ocurre en dos etapas, a través de un estado intermedio que es un complejo covalente enzima-sustrato (figura 1.8.B). En el primer paso, uno de los grupo carboxílicos, que actúa de ácido/base, protona el oxígeno glicosídico, debilitando el enlace. El otro grupo carboxílico actúa de nucleófilo, atacando al carbono anomérico y formando el intermedio covalente enzima-sustrato. En el segundo paso, el residuo ácido/base desprotona la molécula de agua (en el caso de hidrólisis) o la molécula aceptora (transglicosilación), la cuál ataca al carbono anomérico, liberando el azúcar⁵³. Ambos pasos ocurren mediante estados de transición con un carácter ión oxocarbenio considerable. En la formación del estado intermedio covalente, el anillo glicósido covalentemente unido asume una conformación ⁴C₁ de silla estándar⁵⁴.

En las enzimas de retención, la mutación del ácido/base por un grupo metilo (Ala) o por una amida funcionalizada (Asn o Gln) reduce la capacidad de hidrólisis de los sustratos naturales, ya que

se reduce la asistencia de protonación, pero no se ve afectada para compuestos con fuertes grupos salientes. Este es el fundamento de las glicosil sintasas (GS), glicosil hidrolasas con el residuo ácido/base inactivado que son capaces de formar enlaces glicosídicos, partiendo de glicósidos fluorados como substratos donadores. En cuanto al nucleófilo, su papel es clave para las enzimas de retención ya que su mutación causa un efecto mucho más drástico, haciendo imposible la actividad enzimática⁵³.

A) Mecanismo de inversión:



B) Mecanismo de retención:

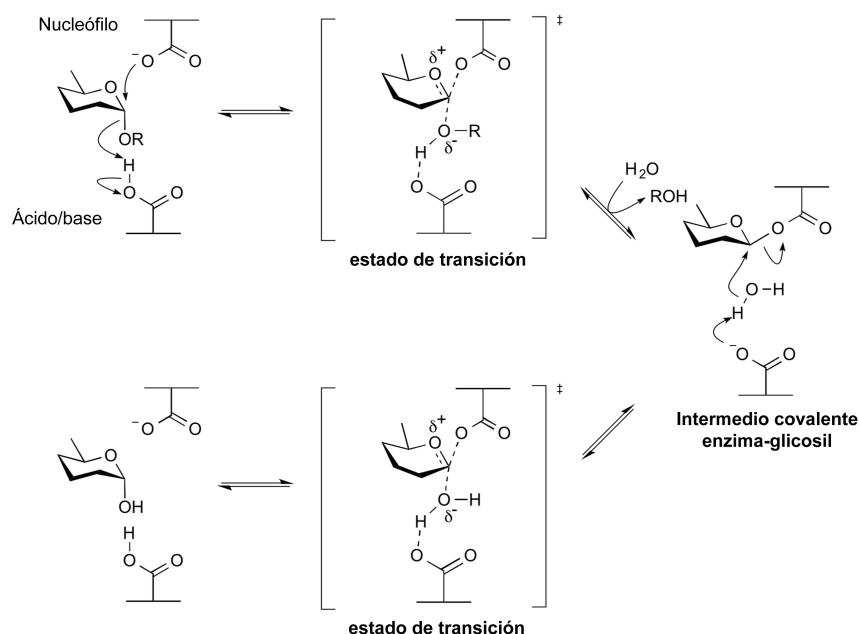


Figura 1.8. Mecanismos de acción de las glicosidasas (A) de inversión o (B) de retención. Imagen adaptada de ⁵³.

Davies G. J., Wilson K. S. y Henrissat B.⁵⁵ definieron la nomenclatura “-n/+n” para denominar los subsitios de unión de las GH, donde la hidrólisis tiene lugar entre los subsitios -1 y +1 (indicado en la figura 1.9). Los subsitios positivos definen el aglicón (grupo saliente tras la hidrólisis) mientras que los sitios de reconocimiento específicos del tipo de azúcar se encuentran en los subsitios negativos. Los estudios estructurales de substratos y análogos unidos a distintas GH han mostrado que los anillos presentan una conformación tipo silla, salvo en el subsitio catalítico -1, que adopta una conformación distorsionada y “pre-activada” hacia el bote. Esta deformación del azúcar en -1 fue asociada con una mejora en la hidrólisis de los glicósidos, ya que la distorsión dirige al substrato en el camino de alcanzar el estado de transición de la reacción, lo que se conoce como “configuración cercana al ataque”, de tal forma que el aglicón se coloca en una posición pseudoaxial (mientras que libre en solución ocupa una posición ecuatorial) y facilita que el nucleófilo ataque al carbono anomérico⁵⁴.

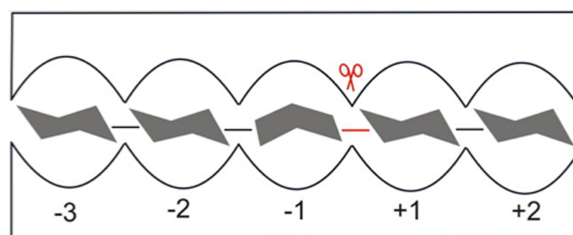


Figura 1.9. Diagrama de los subsitios de unión de las glicosidasas. La hidrólisis ocurre entre los subsitios catalíticos -1 y +1 (señalado en rojo). Modificada de ⁵⁴.

1.4.3) ESTRUCTURA DE LAS GH32

Las enzimas de la familia GH32 degradan fructanos, y también son capaces de sintetizar fructooligosacáridos (FOS). La actividad más característica de esta familia es invertasa, es decir, la degradación de sacarosa, si bien GH32 contiene enzimas capaces de hidrolizar polisacáridos de fructosa, como inulinasas y levanasas, y distintos tipo de fructosiltransferasas.

La primera estructura de la familia GH32 que se resolvió fue la β -fructofuranosidasa de *Thermotoga maritima*⁵⁶. La estructura cristalina reveló que la enzima tiene una arquitectura bimodular, formada por el módulo catalítico β -hélice de 5 láminas (*five-bladed β -propeller*) y un dominio auxiliar β -sándwich.

El dominio catalítico *β -propeller* está compuesto por la repetición de cinco láminas, orientadas de forma radial, enmarcando la cavidad catalítica en el centro de la hélice (figura 1.10). Cada lámina está compuesta por cuatro hebras β antiparalelas, con topología "W", numeradas A, B, C y D desde el interior del dominio hacia la superficie. Los bucles que conectan las láminas entre sí (denominados L₁-L₄) y los bucles que conectan las hebras B y C de cada lámina (denominados giros T₁-T_V) forman el bolsillo catalítico, negativamente cargado. El nucleófilo se encuentra localizado en la primera hebra de la primera lámina, mientras que el residuo ácido/base se encuentra en la primera hebra de la cuarta.

Los dominios *β -propeller* se estabilizan por un mecanismo conocido como "velcro molecular", término que indica que la estructura de hélice se cierra gracias a la interacción de ambos extremos peptídicos del dominio en la misma lámina, enlazando y "sellando" la estructura⁵⁷. El dominio *5-fold β -propeller* caracteriza a los clanes GH-F (familias 43 y 62) y GH-J (familias 32 y 68). Sin embargo, los miembros de las familias GH43 y GH62 presentan un mecanismo de inversión, mientras que las enzimas GH32 y GH68 retienen la configuración del carbono anomérico, por ello se clasifican en clanes diferentes⁵⁸.

Por otro lado, el β -sándwich está compuesto por dos láminas β , de seis hebras β cada una, y su presencia es esencial para la estabilidad de las proteínas GH32. La búsqueda de homólogos estructurales con el servidor DALI (http://ekhidna.biocenter.helsinki.fi/dali_server) señala que el β -sándwich presenta similitudes con las lectinas, que son proteínas que poseen al menos un dominio no-catalítico que une específicamente carbohidratos (mono u oligosacáridos), por lo que podría tener un papel en la unión de los sustratos⁵⁸.

En 2005, Verhaes M. y colaboradores⁵⁹ plantearon un posible papel del β -sándwich en la unión de residuos distales del polímero inulina de la fructan 1-exohidrolasa IIa de *Cichorium intybus*. Sin embargo, no fue hasta 2010, cuando nuestro grupo reportó la β -fructofuranosidasa de *Schwanniomyces occidentalis* (SoFfase), que demostró el papel directo del β -sándwich en la oligomerización enzimática y en la unión del sustrato⁶⁰. Por otra parte, Alberto F. y colaboradores⁵⁶

sugirieron que el β -sándwich podría tener un papel en la oligomerización de la invertasa de *Saccharomyces cerevisiae*. Por tanto, parece que la oligomerización podría ser una peculiaridad de las GH32 de levaduras.

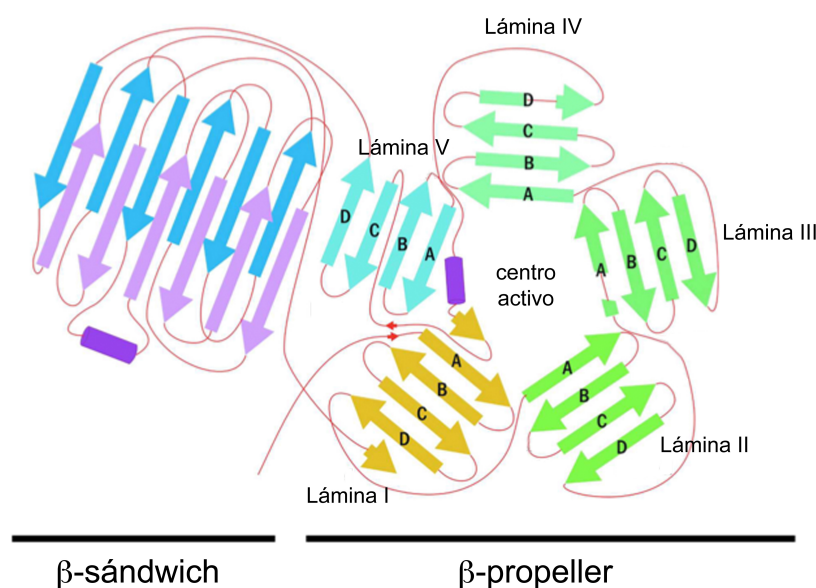


Figura 1.10. Diagrama esquemático de la topología de las enzimas GH32. Las hebras β están representadas como flechas y las α -hélices como cilindros. Modificada de ⁶¹.

CENTRO ACTIVO

Tanto el mecanismo de acción (mecanismo de retención) como los residuos implicados en la catálisis están conservados dentro del clan GH-J. Concretamente, para las enzimas GH32 el nucleófilo es un ácido aspártico (motivo WMNDPNG), mientras que el ácido/base es un ácido glutámico (motivo ECPGL), excepto en la endo-inulinasa de *Aspergillus ficuum*, que presenta un ácido glutámico como nucleófilo⁶². Estos residuos fueron identificados por primera vez en la invertasa de *Saccharomyces cerevisiae*⁶³⁻⁶⁴.

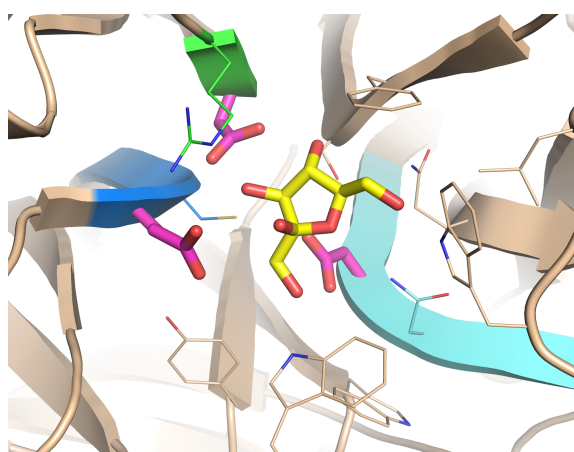


Figura 1.11. Centro activo de una β -fructofuranosidasa GH32. En rosa se resaltan los residuos catalíticos, en azul claro el motivo del nucleófilo, en azul oscuro el motivo del residuo ácido/base y en verde el motivo del ácido aspártico estabilizador. En amarillo se representa una molécula de fructosa en el centro activo de la β -fructofuranosidasa de *Schwanniomyces occidentalis* (código PDB: 3kf3).

Además, hay un tercer residuo catalítico, un ácido aspártico (RDP), que tiene un papel importante como estabilizador del estado de transición. Fue identificado por primera vez en la estructura de la exo-inulinasa de *Aspergillus awamori* en complejo con fructosa⁶⁵, y confirmado en la estructura de la β -fructofuranosidasa de *Thermotoga maritima* en complejo con rafinosa⁶⁶. Además, este residuo reconoce los oxígenos O3 y O4 de la fructosa en -1, por lo que puede ser importante para reconocer los fructofuranósidos, seleccionando negativamente a los fructopiranosidos⁶⁵ (figura 1.11).

1.4.4) ESTRUCTURA DE LAS GH31

La familia GH31 es, junto a la GH13, una de las que más α -glucosidasas contiene. Además del papel esencial que juegan en importantes procesos metabólicos, estas enzimas son útiles para la producción de IMOS. La primera estructura de una enzima GH31 que se dio a conocer fue la α -xilosidasa de *Escherichia coli*⁶⁷. Este trabajo reveló la estructura terciaria característica de esta familia, que está formada por cuatro dominios: un dominio N-terminal β -supersándwich, un dominio catalítico barril (β/α)₈, y dos dominios β -sándwich C-terminales (C-terminal proximal y C-terminal distal) (figura 1.12). Sin embargo, posteriores estudios han mostrado un panorama mucho más complejo, como se verá a continuación.

El dominio N-terminal presenta una topología tipo β -supersándwich, que consta de cuatro láminas β , cada una formada por unas cinco hebras β antiparalelas⁶⁸. Algunas enzimas, como las α -glucosidasas de *Ruminococcus obeum* y *Sulfolobus solfataricus*, presentan un dominio más corto, con sólo 10 hebras β ⁶⁹. Este dominio parece tener un papel en la unión de sustrato así como en el mantenimiento de la integridad del bolsillo catalítico. Presenta un bucle, denominado *N-loop*, que está implicado en la unión de sustrato al centro activo. Como ya se comentará más adelante, está muy conservado entre miembros evolutivamente cercanos y tiene un papel importante en la actividad de la enzima⁶⁸.

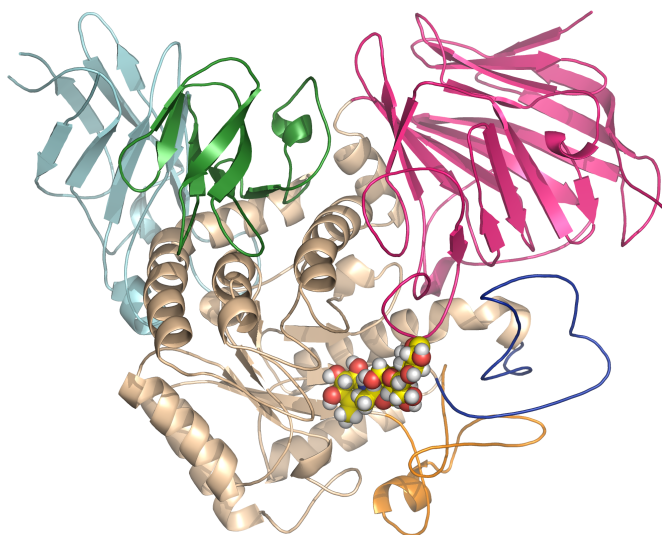


Figura 1.12. Arquitectura molecular de las enzimas de la familia GH31, que constan de cuatro dominios: β -supersándwich N-terminal (rosa), barril (β/α)₈ catalítico (beige), β -sándwich C-terminal proximal (verde) y C-terminal distal (cian). En distinto color se resaltan los bucles implicados en la unión de sustrato (Inserto 1, naranja; Inserto 2, azul oscuro) del dominio catalítico. En esferas se representa una molécula de acarbosa en el centro activo de la enzima α -glucosidasa de *Beta vulgaris* (código PDB: 3wel).

Por otra parte, además de estar involucrado en la unión de sustrato, parece que el dominio N-terminal tiene un papel importante en la oligomerización de algunas enzimas, como es el caso en la hexamerización de la α -xilosidasa de *E. coli*⁶⁷, o en la dimerización de la α -galactosidasa de *Pseudobacter saltans*⁷⁰.

El dominio catalítico es un barril (β/α)₈, compuesto por ocho hebras β y nueve hélices α , muy similar al de la familia GH13 (que también contiene gran cantidad de α -glucosidasas). El dominio catalítico tiene dos largos bucles muy variables, denominados Inserto 1 e Inserto 2, que se sitúan a continuación de las hebras β 3 y β 4, respectivamente. Estos bucles juegan un papel esencial en la actividad, que se comentará en el apartado en el que se describe el centro activo de las enzimas GH31.

El dominio C-terminal proximal está formado por un pequeño β -sándwich central, de ocho hebras β antiparalelas, con dos pequeñas hélices α insertadas en los bucles. Tras el dominio catalítico, es el dominio más conservado⁶⁹. Además, presenta numerosas interacciones hidrofóbicas con el dominio catalítico, por lo que se le puede atribuir un papel estructural y de protección del centro activo. Por otra parte, se ha descrito que este dominio está implicado en la oligomerización de la α -xilosidasa de *E. coli*⁶⁷.

Por último, el dominio C-terminal distal es también un β -sándwich, formado por dos láminas β de unas cinco hebras cada una, con las hebras β 2 y β 6 orientadas en sentido paralelo. Junto con el N-terminal, este dominio es el que presenta menor identidad de secuencia dentro de la familia. De hecho, algunos miembros de la familia presentan una versión más corta, como la α -glucosidasas de *R. obeum*⁶⁹, mientras que otros, como las α -glucosidasas de *Bacteroides thetaiotaomicron* o la sulfoquinosidasa de *E. coli*⁷¹, ni siquiera presentan dominio C-terminal distal. Asimismo, la eliminación de 100 de sus 150 residuos en la α -1-4-glucan-liasas de *Gracilariopsis lemaneiformis* causó un descenso de 10-20 veces en la actividad, sin embargo, estos 100 residuos no fueron necesarios ni para la actividad ni para la expresión de la enzima recombinante⁷².

Ninguno de los dominios C-terminales se sitúa en las proximidades ni parece interactuar con residuos del centro activo. Por tanto, estos dominios posiblemente contribuyen a la estabilidad de la enzima siendo su función estructural, aunque no se puede descartar que puedan tener otro papel aún no descrito⁶⁸.

Además de esta arquitectura molecular fundamental, algunas proteínas de la familia presentan dominios extras. Así, las enzimas humanas (maltasa-glucoamilasa y sucrosa-isomaltasa) muestran un dominio *trefoil* tipo P que precede al dominio N-terminal, cuya función se desconoce. Este dominio de 50 residuos se caracteriza por contener seis cisteínas unidas por puentes disulfuro⁷³. Este tipo de inserciones en el dominio N-terminal son bastante comunes dentro de la familia. Por ejemplo, la sulfoquinosidasa de *E. coli* muestra una larga inserción de unos 50 residuos en el extremo N-terminal, que conforman una larga lámina β de dos hebras antiparalelas, que delimitan uno de los márgenes del centro activo⁷¹. También presentan un fragmento adicional previo al dominio N-terminal las enzimas de *Listeria monocytogenes* y de *Trueperella pyogenes*, que muestran un dominio β -sándwich extra, que no presenta homología con CBM (módulo de unión a carbohidratos, *Carbohydrate binding modules*) u otras proteínas tipo β -sándwich previamente descritas.

Finalmente otras enzimas, como las α -xilosidasas de *Cellvibrio japonicus* y de *Bacteroides ovatus*, tienen un dominio adicional PA14 insertado en el dominio N-terminal⁷⁴⁻⁷⁵. Este dominio adyacente presenta una arquitectura de barril β y se ha descrito en la α -xilosidasa de *C. japonicus* que está implicado en la formación de los subsitios distales +2 y +3 del centro activo⁷⁶.

Todos los miembros del clan GH-D, al que pertenecen las enzimas GH31, presentan el residuo nucleófilo (Asp) en la hebra β 4 mientras que el que actúa como ácido/base (Asp) se encuentra en la hebra β 6. Así, y a pesar de su gran similitud con la familia GH13, la situación del residuo ácido/base es responsable de que GH31 se clasifique en un clan distinto (GH13 presenta el residuo ácido/base en la hebra β 5)^{67, 77}.

Las enzimas GH31 presentan un mecanismo de retención y su centro activo está formado, principalmente, por los extremos de las hebras β del barril (β/α)₈, los bucles del dominio catalítico (principalmente el bucle β 1- α 1, y los Insertos 1 y 2), y el bucle N-*loop* proveniente del dominio N-terminal (figura 1.13).

Al igual que ocurre en otras familias GH, existe poca identidad de secuencia entre los miembros de GH31, aunque presentan zonas altamente conservadas⁷⁸. Una de ellas es la región que contiene el nucleófilo, lo que permite una clasificación de las enzimas GH31 en subfamilias de las cuales, hasta el momento, se han establecido cuatro. Las α -glucosidasas presentan la secuencia consenso WiDMNE (donde i es un residuo variable) y se clasifican en el subgrupo 1, a excepción de algunas α -glucosidasas bacterianas, que presentan la secuencia KTDFGE y se encuentran en el subgrupo 4⁷⁷.

Dentro de los bucles del barril (β/α)₈ que están implicados en el centro activo, el bucle β 1- α 1 determina la especificidad por el tipo de enlace de la enzima. Así, la presencia de un triptófano en este bucle confiere especificidad por enlaces α (1-6), mientras que su mutación por tirosina reduce drásticamente la capacidad de hidrólisis y transglucosilación de este tipo de enlaces, hidrolizando preferentemente enlaces α (1-4)⁶⁹. Sin embargo, en otras enzimas como la α -glucosidasa de *Bacillus sp.*, la especificidad por el tipo de enlace se encuentra en el bucle β 8- α 8⁷⁴.

Por otra parte, el Inserto 1 está relativamente conservado dentro de la familia. Sin embargo, en la maltasa-glucoamilasa humana C-terminal este bucle es más extenso y está implicado en la unión de sustratos largos, conformando los subsitios de unión +2 y +3 de la enzima⁷⁹. Asimismo, las enzimas GH31 que forman o degradan glucosacáridos cíclicos también presentan un bucle de mayor longitud que conforma su subsitio -2.

En cuanto al Inserto 2, su secuencia y su longitud son muy variables. En algunas enzimas GH31 tiene funciones muy importantes en la oligomerización y/o en la determinación de la especificidad de sustrato. Es de destacar el caso de la α -glucosidasa de *Beta vulgaris*, cuya especificidad por sustratos largos reside fundamentalmente en el Inserto 2 que crea subsitios de unión muy alejados del bolsillo catalítico (subsitio +7)⁶⁸.

Por último, el bucle N-*loop* presenta un Asp conservado en todas las enzimas GH31, implicado en la unión del sustrato en el subsitio +1, si bien el resto de la secuencia varía entre los distintos miembros. Aunque hay varias enzimas oligoméricas, el N-*loop* siempre participa en el centro activo del mismo monómero. La secuencia de este bucle se repite en enzimas evolutivamente cercanas. Así, las α -glucosidasas GH31 de plantas, que tienen alta especificidad por sustratos largos, poseen un N-*loop* conservado, que en la α -glucosidasa de *B. vulgaris* conforma los subsitios +2 y +3 de la enzima⁶⁸. Asimismo, los N-*loop* de las α -glucosidasas II de *Mus musculus* y *Chaetomium thermophilus*, que sólo reconocen disacáridos, presentan una fenilalanina que bloquea el centro activo y hace inaccesible los α -glucósidos de más de tres unidades⁸⁰.

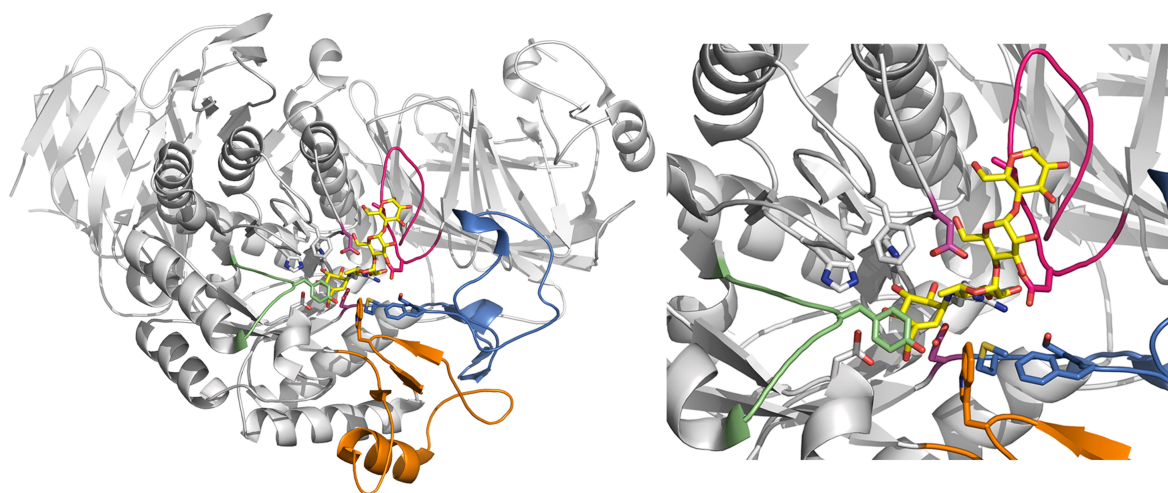


Figura 1.13. Estructura del centro activo de una enzima GH31. Se resaltan los residuos catalíticos (morado), el N-loop (rosa), el bucle de especificidad del tipo de enlace (verde), el Inserto 1 (naranja) y el Inserto 2 (azul). En amarillo se representa una molécula de acarbosa en el centro activo de la maltasa-glucoamilasa humana (Nt-MGAM, código PDB: 2qmj).

1.4.5) ESTRUCTURA DE LAS GH3

La familia 3 de las glicosil hidrolasas (GH3) es una de las familias que mayor número de miembros contiene, con más de once mil secuencias depositadas. Están presentes en bacterias, hongos y plantas, donde realizan una amplia variedad de funciones como la degradación de la biomasa vegetal, remodelado de la pared celular o defensa frente a patógenos. Esta familia contiene principalmente β -glucosidasas, β -xilosidasas, β -N-acetilhexosaminidasas y α -L-arabinofuranosidasas, actuando con un mecanismo catalítico de retención. Algunas presentan una amplia especificidad de sustrato, lo que las convierte en potenciales biocatalizadores de interés en la obtención de nuevos oligosacáridos.

La primera estructura reportada de una enzima GH3 fue la de la β -glucosidasa de *Hordeum vulgare* (cebada), en 1999⁸¹. Este estudio mostró que la enzima estaba compuesta por dos dominios: un barril (β/α)₈ y un dominio sándwich (β/α)₆. La siguiente estructura resuelta fue la de la β -N-acetilhexosaminidasa de *Vibrio cholerae*, que mostró que esta enzima bacteriana estaba compuesta únicamente por el dominio barril (β/α)₈⁸². Estudios posteriores realizados con la β -glucosidasa de *Thermotoga neapolitana* mostraron una estructura compuesta por tres dominios; además de los dos observados anteriormente en la enzima de planta, esta enzima contiene en el extremo C-terminal un dominio adicional denominado FnIII⁸³.

En los últimos años han aparecido numerosas estructuras de enzimas GH3, que muestran una gran variedad en la composición y disposición de hasta cuatro o cinco dominios distintos. En general, las β -glucosidasas y la β -xilosidasa GH3 tienen una composición de dominios básica que corresponde a lo observado en la β -glucosidasa de *T. neapolitana* (figura 1.14).

El dominio N-terminal presenta una topología de barril (β/α)₈, formado por ocho láminas β paralelas rodeadas de ocho hélices α , que se alternan en secuencia⁸¹. La estructura de la β -glucosidasa de *T. neapolitana* reveló una desviación de la forma canónica de la topología del barril (β/α)₈, ya que carece de su primera hélice α , lo que se traduce en que la hebra β 2 es antiparalela a las otras siete hebras⁸³. En realidad este dominio no canónico se define mejor como un barril $\beta\beta(\beta/\alpha)$ ₆⁸⁴. Esta variación en la topología del barril es inusual en las bases de datos CATH y SCOP⁸³, sin embargo, otras enzimas de la familia GH3, resueltas posteriormente, también la presentan. De

hecho, la mayoría de β -glucosidasas resueltas presentan esta inversión de la hebra $\beta 2$, excepto las β -glucosidasas bacterianas de *Listeria innocua* o *Bacteroides ovatus*, y la aislada de compost.

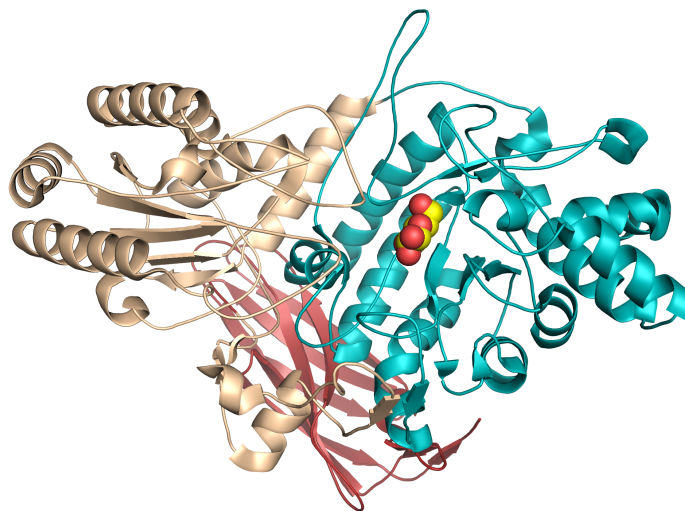


Figura 1.14. Arquitectura molecular de las β -glucosidasas y β -xilosidasas de la familia GH3, compuesta por tres dominios: dominio catalítico barril $(\beta/\alpha)_8$ (azul), dominio sándwich $(\beta/\alpha)_6$ (beige) y dominio FnIII (rojo). En esferas se representa una molécula de glucosa en el centro activo de la β -glucosidasa de *Thermotoga neapolitana* (código PDB: 2x41).

El segundo dominio muestra una topología de sándwich $(\beta/\alpha)_6$, que consiste en una lámina β de seis hebras (cinco de las cuales son paralelas), además de tres hélices α a ambos lados de la lámina. Este dominio está conectado al barril $(\beta/\alpha)_8$ por una hélice α ⁸¹. Este dominio está menos conservado que el catalítico, con variación tanto en la longitud como en el número de hebras y hélices. Asimismo, la longitud de los bucles que las conectan es muy variable. Por ejemplo, en la β -glucosidasa de *T. neapolitana* la hebra $\beta 3$ está sustituida por una hélice α junto a un largo bucle que se proyecta hacia el centro activo y podría estar implicado en la unión de sustratos⁸³.

Por último, las β -glucosidasas y la β -xilosidasa GH3 presentan en el extremo C-terminal de sus secuencias el dominio FnIII (Dominio fibronectina tipo III, *Fibronectin-like type III domain*), cuya función no se conoce. La fibronectina, donde se identificó por primera vez este dominio, contiene 16 copias de este motivo estructural, que presenta una topología de β -sándwich y está formado por dos láminas de 3-4 hebras β cada una⁸⁵. Por su localización, parece que el dominio FnIII no tiene un papel directo en la actividad de la enzima. Experimentos con quimeras donde se translocó el dominio FnIII de la enzima *Cellvibrio gilvus* (mesófila) a la enzima de *T. maritima* (termófila) mostraron que los valores de actividad de la quimera se parecían a los de *T. maritima*, por lo que este dominio no afecta a la actividad de la enzima⁸⁶. No obstante, el dominio FnIII podría estar implicado en el anclaje de la enzima a polímeros o cumplir otras funciones de anclaje celular, o incluso tener un papel en la termoestabilidad⁸³.

Excepcionalmente, algunas enzimas GH3 presentan dominios extra. Las β -glucosidasas de *Kluyveromyces marxianus*⁸⁴ y *Streptomyces venezuelae*⁸⁷ presentan un dominio PA14 extra, insertado en la secuencia del dominio sándwich $(\beta/\alpha)_6$. Este dominio consiste en un barril β formado por dos láminas de 5-6 hebras β cada una. Los extremos N- y C-terminal de este dominio se encuentran muy cercanos en el espacio, lo que facilita su inserción dentro de otros dominios, sin obstaculizar su estructura⁸⁸. En la β -glucosidasa de *K. marxianus* se demostró, por primera vez, la interacción entre el dominio PA14 y un carbohidrato. Así, este dominio está implicado en la formación del subsitio +1 el centro activo de la enzima⁸⁴.

Otra enzima GH3 reportada con dominios adicionales es la exo-1,3/1,4- β -glucanasa de *Pseudoalteromonas* sp. BB1⁸⁹. Además del esqueleto básico que contiene el barril (β/α)₈ y el sándwich (β/α)₆, presenta otro dominio β -sándwich C-terminal, no descrito con anterioridad. Un análisis comparativo utilizando el servidor DALI revela que su homólogo estructural más cercano es un CBM30, aunque no presenta las características típicas de un dominio de unión a carbohidratos. Como se verá posteriormente, una de las enzimas incluidas en esta tesis, perteneciente a la familia GH3, presenta un dominio cuyos homólogos más próximos son también dominios CBM y tampoco presenta motivos obvios de unión a carbohidratos. Por tanto, quedan aún muchos aspectos por entender para conocer las intrincadas bases moleculares de la actividad de esta compleja familia de proteínas.

CENTRO ACTIVO

La mayoría de las β -N-acetilhexosaminidasas (NagZ) de la familia están formadas únicamente por el dominio catalítico, si bien hay otras, como la NagZ de *Bacillus subtilis*⁹⁰, que contienen también el dominio sándwich (β/α)₆. En cualquier caso, los residuos catalíticos de todas las NagZ están presentes únicamente en el barril (β/α)₈.

Por el contrario, los estudios realizados con la β -glucosidasa de *H. vulgare* revelaron que el centro activo se localiza entre ambos dominios⁸¹. La contribución de diferentes dominios en el suministro de los residuos catalíticos es una característica muy inusual dentro de las GH, pero este fenómeno se conserva en los miembros GH3 que tienen dos o más dominios (no NagZ). Así, el nucleófilo (Asp) se sitúa en la hebra β 7 del barril (β/α)₈, mientras que el residuo ácido/base (Glu) se encuentra en un largo bucle que se extiende desde el dominio sándwich (β/α)₆⁹¹.

Las estructuras de β -glucosidasas y β -xilosidasas GH3 conocidas hasta la fecha muestran que, si bien el sitio -1 está generalmente conservado, existen importantes diferencias en los residuos que configuran el centro activo. Así, las β -glucosidasas presentan un triptófano que orienta el primer anillo del azúcar (sitio +1), mientras que este residuo aromático está ausente en la β -xilosidasa. El subsitio +1 es muy variable entre las distintas enzimas, aunque en la mayoría de las β -glucosidasas uno o dos residuos aromáticos forman una plataforma hidrofóbica que orienta el segundo anillo del carbohidrato (figura 1.15).

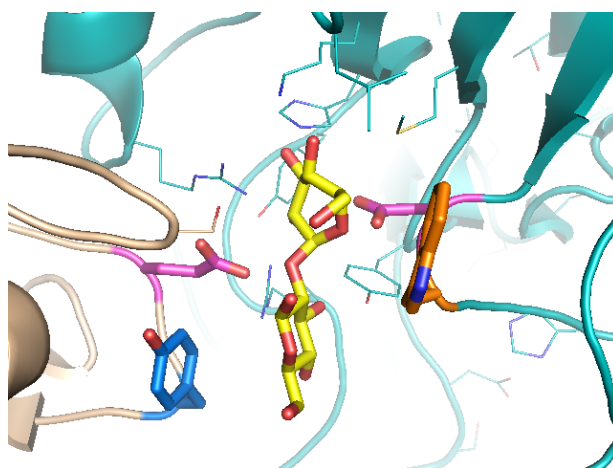


Figura 1.15. Centro activo de una β -glucosidasa GH3. En rosa se resaltan los residuos catalíticos, en naranja el residuo aromático que orienta el anillo de glucosa en el subsitio -1, y en azul el residuo aromático que forma la plataforma hidrofóbica del subsitio +1 de la enzima. En amarillo se representa una molécula de soforosa (Glu- β (1-2)-Glu) en el centro activo de la β -glucosidasa de *Rasamsonia emersonii* (código PDB: 4d0j y 4zo8).

1.5) LA BIOTECNOLOGÍA ENZIMÁTICA

Se considera Biotecnología a toda “aplicación tecnológica que utilice sistemas biológicos, organismos vivos o sus derivados para la creación o modificación de productos o procesos para usos específicos”⁹².

El trabajo de la Biotecnología es multidisciplinar e incluye herramientas tomadas de la Química, la Bioquímica y la Biología Molecular. Igual de diversas son sus aplicaciones, que incluyen campos como la Medicina, la Agricultura, la Alimentación o el Medio Ambiente. Por ello, la Biotecnología se ha clasificado por colores, que representan los campos de aplicación. Así, la Biotecnología Roja se centra en las aplicaciones medicas y farmacéuticas, como la producción de vacunas y antibióticos, el descubrimiento de nuevos fármacos, las terapias regenerativas o la construcción de órganos artificiales. Mientras, la Biotecnología Blanca se ocupa de la biocatálisis y los procesos industriales a gran escala, como la producción de biodiesel, la síntesis de fármacos o la producción de suplementos alimentarios⁹³.

La Biocatálisis se ha establecido como una técnica más eficiente que la síntesis química para la producción de diversos compuestos orgánicos, además de ser una tecnología respetuosa con el medio ambiente. Actualmente, incluye el uso de microorganismos o de enzimas, tanto para laboratorio como para escala industrial. Esta técnica ha sufrido una gran transformación desde las primeras aplicaciones de extractos naturales para la síntesis de compuestos químicos hace unos 100 años. La Ingeniería Genética, la tecnología del control y la transferencia del ADN de un organismo a otro, permitió aislar las enzimas del organismo origen. Pero no solo eso, sino que también permitía modificar dichas proteínas, consiguiendo funciones mejoradas o incluso nuevas. Posteriormente, la aparición de las primeras estructuras de proteínas, que mostraron los mecanismos moleculares de acción de las enzimas, fueron un pilar importante para el desarrollo de la Ingeniería de Proteínas hace 30 años⁹⁴.

Así, al principio el proceso se diseñaba limitado por las características enzimáticas. En la actualidad, sin embargo, es la enzima la que se ajusta a las necesidades y especificaciones del proceso catalítico⁹⁵. En este nuevo proceso de diseño, el conocimiento estructural de la enzima es fundamental para el diseño racional de mutaciones.

APORTACIÓN DE LA BIOLOGÍA ESTRUCTURAL A LA BIOTECNOLOGÍA ENZIMÁTICA

La Biotecnología Enzimática ha permitido la síntesis regio-selectiva de compuestos naturales, así como la producción de nuevos compuestos químicos, con múltiples aplicaciones farmacéuticas. Sin embargo, ha sido durante los últimos 10-15 años cuando la Biotecnología Enzimática está sufriendo su mayor evolución, gracias al gran aumento en el número de estructuras tridimensionales resueltas, que facilita la Ingeniería de Proteínas. El número de estructuras depositadas en la base de datos de proteínas RCSB (PDB, www.pdb.org) ha aumentado en la última década y media en un 750%. A fecha de escritura de esta tesis, el RCSB PDB contiene más de 110.000 estructuras de proteínas, el 90% de las cuales han sido resueltas por Cristalografía de rayos X.

Esta enorme base de datos de estructuras facilita tanto el diseño racional como la evolución dirigida, ya que el alineamiento de estructuras de proteínas relacionadas ayuda a identificar las diferencias y similitudes para un diseño más preciso de librerías de mutantes, lo que se conoce como diseño semi-racional de proteínas. Este método híbrido combina las ventajas del diseño racional (selección de puntos de mutación) y las herramientas de la evolución dirigida (sustituciones de

residuos al azar mediante mutagénesis por saturación). Además, otras herramientas bioinformáticas, como la generación de modelos estructurales bioinformáticos, aportan herramientas muy útiles para la Biotecnología Enzimática⁹⁴.

Por tanto, este trabajo pretende contribuir al avance en el diseño más eficaz de biocatalizadores mediante el análisis en profundidad de las bases moleculares de la especificidad de cuatro enzimas productoras de oligosacáridos bioactivos.

Más información sobre la metodología propia de la Cristalografía y la Cristalización de proteínas se encuentra en el ANEXO II. Así mismo, también se ha incluido un apartado sobre otras técnicas de Biología Estructural utilizadas en esta tesis (SAXS y modelado bioinformático) en el ANEXO III.

2. OBJETIVOS

El objetivo general del trabajo es avanzar en el esclarecimiento de los mecanismos de reconocimiento proteína-carbohidrato que modulan la especificidad de glicosidasas capaces de producir oligosacáridos. Este conocimiento es la base fundamental para realizar, de manera dirigida y eficaz, la ingeniería molecular de enzimas con fines biotecnológicos. El objetivo último es producir oligosacáridos con propiedades prebióticas, diseñados específicamente para manipular la microbiota intestinal del individuo, lo que abre la puerta a una futura medicina preventiva personalizada.

Los objetivos concretos del trabajo son:

1. Estudio cristalográfico de la invertasa de *Saccharomyces cerevisiae*, una enzima productora de FOS. Caracterización estructural del estado octamérico de la enzima. Análisis de los mecanismos moleculares de la especificidad de sustrato y caracterización de su actividad. Estudio del papel del β -sándwich en la alta oligomerización de esta enzima.
2. Cristalización y resolución de la estructura del mutante inactivo de la β -fructofuranosidasa de *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*), libre y en complejo con distintos fructooligosacáridos. Estudio de los elementos moleculares implicados en la especificidad de producción de FOS y diseño de mutantes. Caracterización de su actividad y búsqueda de nuevos productos de transglicosilación.
3. Predicción estructural de la α -glucosidasa de *Xanthophyllomyces dendrorhous*. Caracterización de su actividad y análisis de los posibles mecanismos moleculares implicados en la especificidad de producción de IMOS.
4. Cristalización de una β -glicosidasa de metagenoma de rumen. Resolución estructural de la enzima, libre y en complejo con distintos β -monosacáridos. Análisis de los elementos estructurales implicados en la amplia especificidad hidrolítica de la enzima. Estudio del papel del dominio extra C-terminal en la estabilidad enzimática, así como en la actividad y en la especificidad sobre el sustrato.

3. ENZIMAS PRODUCTORAS DE FOS:

Invertasa de *S. cerevisiae* y

β -Fructofuranosidasa de

X. dendrorhous

3.1) INTRODUCCIÓN

Los fructanos son polímeros de fructosa biosintetizados a partir de sacarosa por muchas especies de microorganismos y plantas. Si bien su principal papel es el almacenamiento de carbohidratos y la reserva de energía, los fructanos tienen otras importantes funciones fisiológicas. En las plantas, la sacarosa juega un papel vital en la señalización de las necesidades energéticas del organismo, regulando su crecimiento o inhibiendo la fotosíntesis. Así, el equilibrio fructanos-sacarosa marca las necesidades metabólicas de la planta. Además, algunos fructanos, como los graminanos, intervienen en la protección de las plantas frente a condiciones adversas del entorno⁹⁶. Asimismo, los fructanos están implicados en la relación entre los distintos reinos. En las bacterias, los fructanos actúan a modo de barrera física, que reduce el estrés al tiempo que asisten en la colonización microbiana⁹⁷. Esta simbiosis también se podría observar entre las levaduras que viven en la materia vegetal. Es por este papel fundamental en el metabolismo de los microorganismos y las plantas que el estudio estructural de las enzimas que modifican fructanos es de gran relevancia y tiene importantes implicaciones en procesos biológicos esenciales.

La degradación de fructanos y la producción de fructooligosacáridos (FOS) es llevada a cabo por enzimas pertenecientes a las familias de glicosil hidrolasas GH32 y GH68⁹⁸. A pesar de su baja similitud de secuencia (<15% identidad), estas dos familias comparten un mismo plegamiento del centro activo y el mismo mecanismo de acción, por ello están incluidas dentro del mismo clan GH-J. Las GH68, formadas sólo por el dominio catalítico, agrupan las fructan hidrolasas de bacteria, tanto las levansucrasas y las inulosucrasas, como las invertasas⁵⁸. Sin embargo, las enzimas de la familia GH32, que agrupan enzimas de eucariotas y bacterias con actividad fructosil hidrolasa o transferasa, presentan, además del dominio catalítico, un dominio adyacente β -sándwich. Este dominio C-terminal está directamente implicado en la oligomerización de la β -fructofuranosidasa de *S. occidentalis* (SoFfase), la primera estructura dimérica resuelta de la familia GH32⁶⁰. La resolución de esta estructura mostró además que el β -sándwich tiene un papel fundamental en el reconocimiento del sustrato.

En cuanto a la síntesis industrial de FOS, estos prebióticos se pueden producir por la acción de fructosil transferasas (FT) o fructosil hidrolasas (FH). Las FT producen una mezcla de fructooligosacáridos con distinto grado de polimerización (de 3-6 unidades, pudiendo llegar hasta 10-15 unidades)⁹⁹. Habitualmente, presentan preferencia por un tipo de sustrato, aunque algunas pueden usar como aceptores fructanos tipo inulina o tipo levano, como la fructosil transferasa de *Pachysandra terminalis*¹⁰⁰. Análogamente, el producto que generan suele contener sólo un tipo de enlace.

En condiciones de exceso de sustrato, las FH también presentan actividad transfructosilasa. La ventaja es que generan un producto menos heterogéneo, y con un menor grado de polimerización, siendo los trisacáridos los productos principales⁵⁸. En cuanto al tipo de enlace generado, se han descrito FH con capacidad de generar FOS con enlaces β (2-1) o β (2-6)³⁴. No obstante, la ingeniería

enzimática ha permitido el diseño de mutantes que mejoran la especificidad de la actividad transferasa produciendo casi exclusivamente un tipo de FOS¹⁰¹.

Hasta la fecha de escritura de este manuscrito, se han resuelto 14 estructuras tridimensionales de la familia GH32, (www.cazy.org/GH32_structure.html) todas por cristalografía de rayos X, siendo las estructuras presentadas en este trabajo las dos últimas resueltas.

Three-Dimensional Structure of *Saccharomyces* Invertase. Role of a Non-Catalytic Domain in Oligomerization and Substrate Specificity.

M. Ángela Sainz-Polo‡, Mercedes Ramírez-Escudero‡, Álvaro Lafraya, Beatriz González, Julia Marín-Navarro, Julio Polaina and Julia Sanz-Aparicio.

The Journal of Biological Chemistry, Vol. 288, No. 14, pp. 9755–9766,

5 de Abril del 2013

‡ Ambos autores contribuyeron equitativamente a este trabajo.

DOI: 10.1074/jbc.M112.446435

Código de acceso del PDB: 4EQV

Interés de la investigación:

La invertasa de *Saccharomyces cerevisiae* (ScInv) es la enzima que degrada la sacarosa o azúcar común, ampliamente utilizada en la industria. Además, fue una de las primeras enzimas aisladas, por lo que ha constituido el modelo clásico usado en los laboratorios de bioquímica conduciendo, por ejemplo, a la formulación del modelo cinético de Michaelis-Menten en 1913. A pesar de sus numerosas aplicaciones y su importancia histórica, su estructura era desconocida.

ScInv es capaz de producir FOS con enlaces $\beta(2-6)$, los cuales son de interés creciente dado que se ha descrito una mejor actividad prebiótica, en comparación a los clásicos $\beta(2-1)$ -FOS. Por ello, conocer su estructura en detalle es esencial para caracterizar los mecanismos de producción de ^6F -FOS.

Por otra parte, diversos estudios mostraron la existencia de distintos estados oligoméricos en las formas intra/extra celular de ScInv, lo cuál fue corroborado por experimentos de microscopía electrónica. Por tanto, la resolución estructural de ScInv puede explicar también los mecanismos implicados en la oligomerización de las enzimas de levadura de la familia GH32.

Los objetivos de este trabajo fueron:

- Establecer las condiciones óptimas de cristalización y difracción de ScInv.
- Resolver la estructura tridimensional de ScInv.
- Interpretar los datos estructurales y bioquímicos obtenidos, analizar los complejos realizados mediante *docking*, e investigar la funcionalidad e implicación de los distintos bucles del centro activo en el mecanismo de acción de ScInv.
- Estudiar las bases moleculares implicadas en la oligomerización.

- Investigar el papel que tiene el β -sándwich en la oligomerización y en la especificidad de la actividad de ScInv.
- Interpretar las formas octaméricas intracelulares y extracelulares de ScInv.

Las conclusiones de esta investigación son:

- El monómero y el dímero de ScInv son muy similares a los observados previamente en la β -fructofuranosidasa de *Schwanniomyces occidentalis* (SoFfase). Sin embargo, la diferente secuencia en la hebra β 1- β 2 del dominio β -sándwich es responsable de la existencia de estados de oligomerización superiores en ScInv.
- ScInv genera dos tipos de octámeros, uno extracelular abierto, con forma de herradura, y otro intracelular cerrado. El octámero se genera través de la extensión de la lámina beta del dominio β -sándwich.
- En el octámero intracelular de ScInv, mejor descrito como tetrámero de dímeros, existen dos tipos de dímeros: abierto y cerrado.
- A pesar de que ScInv y SoFfase comparten una alta identidad de secuencia y presentan dímeros semejantes, sus perfiles de actividad son distintos. Mientras que SoFfase exhibe actividad invertasa e inulinasa, la configuración octamérica de ScInv impide el reconocimiento de fructooligosacáridos con más de 3-4 unidades, presentando por tanto actividad esencialmente invertasa.

La autora de esta tesis ha realizado:

- Análisis de los estados de oligomerización de ScInv bajo distintos tratamientos (calor, urea, alta salinidad, dilución de la proteína y variación del pH), y su posterior examen en geles PAGE para analizar la distribución de los posibles estados oligoméricos en cada muestra.
- Análisis de la actividad en todos los estados oligoméricos mediante zimograma.
- Estudio del papel del β -sándwich en la determinación del estado de oligomerización. Análisis secuencial de distintas invertasas y β -fructofuranosidasas, y de los patrones de oligomerización de las enzimas de levadura (SoFfase y ScInv).

Three-dimensional Structure of *Saccharomyces* Invertase

ROLE OF A NON-CATALYTIC DOMAIN IN OLIGOMERIZATION AND SUBSTRATE SPECIFICITY*

Received for publication, December 20, 2012, and in revised form, February 20, 2013. Published, JBC Papers in Press, February 21, 2013, DOI 10.1074/jbc.M112.446435

M. Angela Sainz-Polo^{†1,2}, Mercedes Ramírez-Escudero^{†1}, Alvaro Lafraya^{§3}, Beatriz González[‡], Julia Marín-Navarro[§], Julio Polaina[§], and Julia Sanz-Aparicio^{†4}

From the [‡]Departamento de Cristalografía y Biología Estructural, Instituto de Química-Física "Rocasolano," Consejo Superior de Investigaciones Científicas (CSIC), Serrano 119, 28006 Madrid and the [§]Instituto de Agroquímica y Tecnología de Alimentos, CSIC, 46980 Paterna, Valencia, Spain

Background: Invertase is a fundamental enzyme for sugar metabolism in yeast and a classical model in early biochemical studies.

Results: Invertase shows an unusual octameric quaternary structure composed of two types of dimers.

Conclusion: A peculiar pattern of monomer assembly through non-catalytic domain interactions determines invertase specificity.

Significance: Unraveling the structural features that rule enzyme modularity casts new light on protein-carbohydrate recognition.

Invertase is an enzyme that is widely distributed among plants and microorganisms and that catalyzes the hydrolysis of the disaccharide sucrose into glucose and fructose. Despite the important physiological role of *Saccharomyces* invertase (SInv) and the historical relevance of this enzyme as a model in early biochemical studies, its structure had not yet been solved. We report here the crystal structure of recombinant SInv at 3.3 Å resolution showing that the enzyme folds into the catalytic β -propeller and β -sandwich domains characteristic of GH32 enzymes. However, SInv displays an unusual quaternary structure. Monomers associate in two different kinds of dimers, which are in turn assembled into an octamer, best described as a tetramer of dimers. Dimerization plays a determinant role in substrate specificity because this assembly sets steric constraints that limit the access to the active site of oligosaccharides of more than four units. Comparative analysis of GH32 enzymes showed that formation of the SInv octamer occurs through a β -sheet extension that seems unique to this enzyme. Interaction between dimers is determined by a short amino acid sequence at the beginning of the β -sandwich domain. Our results highlight the role of the non-catalytic domain in fine-tuning substrate specificity and thus supplement our knowledge of the activity of this important family of enzymes. In turn, this gives a deeper insight into the structural features that rule modularity and protein-carbohydrate recognition.

Invertase (EC 3.2.1.26; β -fructofuranosidase) catalyzes the hydrolysis of the disaccharide sucrose (table sugar) into glucose and fructose and is a major enzyme present in plants and microorganisms. Because the yeast *Saccharomyces* was one of the preferred materials in early biochemical studies, yeast invertase became one of the classical model enzymes. Mitscherlich described in 1842 the existence in yeast of a substance capable of inverting dextrorotatory cane sugar into a levorotatory sugar that was identified in 1847 by Dubrunfaut as a mixture of glucose and fructose. In 1860, for the first time, Berthelot carried out the isolation of invertase (see Ref. 1). Some years later, the whole theory of enzyme kinetics was based on experimental results obtained with yeast invertase (2). Because invertase was found to be intimately associated with the yeast cells and because its purification required the preparation of yeast extracts, it was considered to be an intracellular enzyme. However, de la Fuente and Sols (3) showed that the enzyme is secreted by yeast cells and that the hydrolysis of sucrose is extracellular. Subsequently, it was discovered that the yeast produces two types of invertase: a heavily glycosylated secreted form and a non-glycosylated intracellular form (4, 5). Yeast invertase is encoded by a family of repeated *SUC* genes (6–9). The enzyme is normally secreted by yeast as a heavily glycosylated octameric protein. The large mass of the protein leads to it being trapped in the cell wall (10, 11). Both the non-glycosylated cytoplasmic form and the secreted form of invertase are encoded by the same gene (8). These two forms are transcribed as two mRNAs of different length, which are translated into polypeptides of different size. In addition, the longer one encodes the signal peptide needed for secretion (12).

On the basis of sequence similarity, invertase is classified within family 32 of the glycoside hydrolases (13). In addition to invertases, this family (designated GH32) includes inulinases and levanases, involved in the hydrolysis of fructose-containing polysaccharides, and also transglycosylases with fructose transferase activity. GH32 enzymes act by a retaining mechanism in

*This work was supported in part by Grants BIO2010-20508-C04-02 and BIO2010-20508-C04-03 from the Spanish Ministry of Education and Science.

The atomic coordinates and structure factors (code 4EQV) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Both authors contributed equally to this work.

² Supported by a Junta para la Ampliación de Estudios (JAE) predoctoral fellowship from the Consejo Superior de Investigaciones Científicas.

³ Present address: Instituto de Catálisis y Petroleoquímica, CSIC, E-28049 Madrid, Spain.

⁴ To whom correspondence should be addressed. Tel.: 34-91-561-9400; Fax: 34-91-564-2431; E-mail: xjulia@iqfr.csic.es.

Structure of *Saccharomyces Invertase*

which an aspartate located close to the N terminus acts as the catalytic nucleophile and a glutamate acts as the general acid/base catalyst. The reaction proceeds through attachment of the aspartate nucleophile to a fructosyl unit of the donor substrate. The fructosyl is subsequently released by hydrolysis or transferred to an acceptor sugar substrate (transglycosylation). In recent years, the crystallographic structure of several GH32 enzymes from bacteria (14–16) and eukaryotes (17–21) have been reported. A close phylogenetic relative to *Saccharomyces invertase* (SInv)⁵ is an inulinase from *Schwanniomyces occidentalis* (22). The characteristic structural feature of GH32 enzymes, shared by GH68 enzymes included in the same GH-J clan, is a 5-fold β -propeller catalytic domain consisting of five blades (each composed of four antiparallel β -strands with a “W” topology), which surround a central negatively charged active site cavity. The GH32 family differs from the GH68 family by an additional β -sandwich domain appended to the catalytic domain.

Besides its historical importance in the development of biochemistry, SInv has extensive industrial applications. It is one of the most widely used enzymes in confectionary to make liquid centers in candy making, and it is also used during fermentation of cane molasses into ethanol. A new potential application is the synthesis of prebiotic fructo-oligosaccharides used in functional foods and pharmaceutical formulations (23). The use of prebiotics to orchestrate the gut microbiota composition, with the associated benefit to human health, is an emerging issue of the utmost biotechnological interest (24).

In this work, we report the three-dimensional structure of SInv, produced in *Escherichia coli*, by expression of the *SUC2* coding sequence without the 5'-end of the gene corresponding to the secretion signal peptide. Our results reveal the structural basis of the unique oligomerization pattern observed in SInv and provide key factors to understanding the enzymatic activity and specificity of this important enzyme.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification—The gene encoding SInv was expressed in *E. coli* Rosetta 2 cells carrying plasmid pQE-SUC2. The resulting His-tagged protein was purified by nickel affinity chromatography as described previously (23). For crystallization, the eluted fractions containing Suc2 (>95% purity according to SDS-PAGE analysis and Coomassie Blue staining) were dialyzed (1:10,000) against 0.05 M Tris-HCl buffer (pH 7) and concentrated using 20-kDa cutoff membrane ultrafiltration (Pierce). For activity assays, SInv was dialyzed (1:10,000) against 0.05 M phosphate buffer (pH 7) and 150 mM NaCl.

Crystallization and Data Collection—Crystallization of SInv (10 mg/ml in 0.05 M Tris-HCl (pH 7)) was performed on Cryschem sitting drop plates (Hampton Research) at 18 °C. Small bars grew in 3% PEG 3350, 5% 2-methyl-2,4-pentanediol, 0.6 M magnesium formate, 1 mM tris(2-carboxyethyl)phosphine, 0.33 M guanidinium chloride, and 0.1 Bistris (pH 6.5)

TABLE 1

Crystallographic statistics of SInv

Values in parentheses are for the high resolution shell. r.m.s.d., root mean square deviation.

Crystal data	
Space group	P3 ₁ 21
Unit cell parameters (Å)	
<i>a</i>	268.6
<i>b</i>	268.6
<i>c</i>	224.4
Data collection	
ESRF Beamline	ID23-1
Temperature (K)	100
Wavelength (Å)	0.979
Resolution (Å)	56.31–3.30 (3.48–3.30)
Data processing	
Total reflections	873,783 (134,865)
Unique reflections	137,870 (17,424)
Multiplicity	6.4 (5.0)
Completeness (%)	98.5 (92.6)
Mean <i>I</i> / σ (<i>I</i>)	8.1 (2.0)
<i>R</i> _{merge} (%) ^a	20.5 (55.6)
<i>R</i> _{pim} (%) ^b	9.7 (37.7)
Molecules/asymmetric unit	8
Matthews coefficient (Å ³ Da ^{−1})	24.9
Solvent content (%)	75
Refinement	
<i>R</i> _{work} / <i>R</i> _{free} (%) ^c	22.2/23.9
Mean <i>B</i> -factors	
Peptide	51.2
Water	20.3
No. of atoms	
Protein	33,016
Water molecules	32
Ramachandran (32) (%)	
Favored	89.1
Outliers	2.5
r.m.s.d.	
Bonds (Å)	0.008
Angles	1.18°
Protein Data Bank code	4EQV

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement of reflection *hkl*, and $\langle I(hkl) \rangle$ is the weighted mean of all measurements.

^b $R_{\text{pim}} = \sum_{hkl} (1/(N-1))^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where *N* is the redundancy for the *hkl* reflection.

^c $R_{\text{work}}/R_{\text{free}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$, where F_c and F_o are the calculated and observed structure factor amplitudes of reflection *hkl* for the working/free (5%) set, respectively.

(protein/precipitant drop ratio of 2:1) by vapor diffusion at room temperature. The crystals diffracted to 3.3 Å resolution. Many attempts to improve resolution at low temperature or using other techniques such as streak and microseeding or using agarose (2–3%) in the crystallization drops were unsuccessful. For more details about the crystallization behavior of this SInv, see Ref. 25.

Crystals of SInv belonged to space group P3₁21, with eight molecules in the asymmetric unit and 75% solvent content within the unit cell. For data collection, native crystals were transferred to cryoprotectant solutions consisting of mother liquor plus 25% (v/v) glycerol before being cooled to 100 K in liquid nitrogen. Diffraction data were collected using synchrotron radiation at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) on beamline ID23-1. Diffraction images were processed with iMOSFLM (26) and merged using the CCP4 package (27). A summary of data collection and data reduction statistics is provided in Table 1.

Structure Solution and Refinement—The structure of SInv was solved by molecular replacement using the MOLREP program (27). The structure of *S. occidentalis* β -fructofuranosidase (SoFfase; Protein Data Bank code 3KF3) was used to pre-

⁵ The abbreviations used are: SInv, *Saccharomyces invertase*; Bistris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; SoFfase, *S. occidentalis* β -fructofuranosidase; NCS, non-crystallographic symmetry.

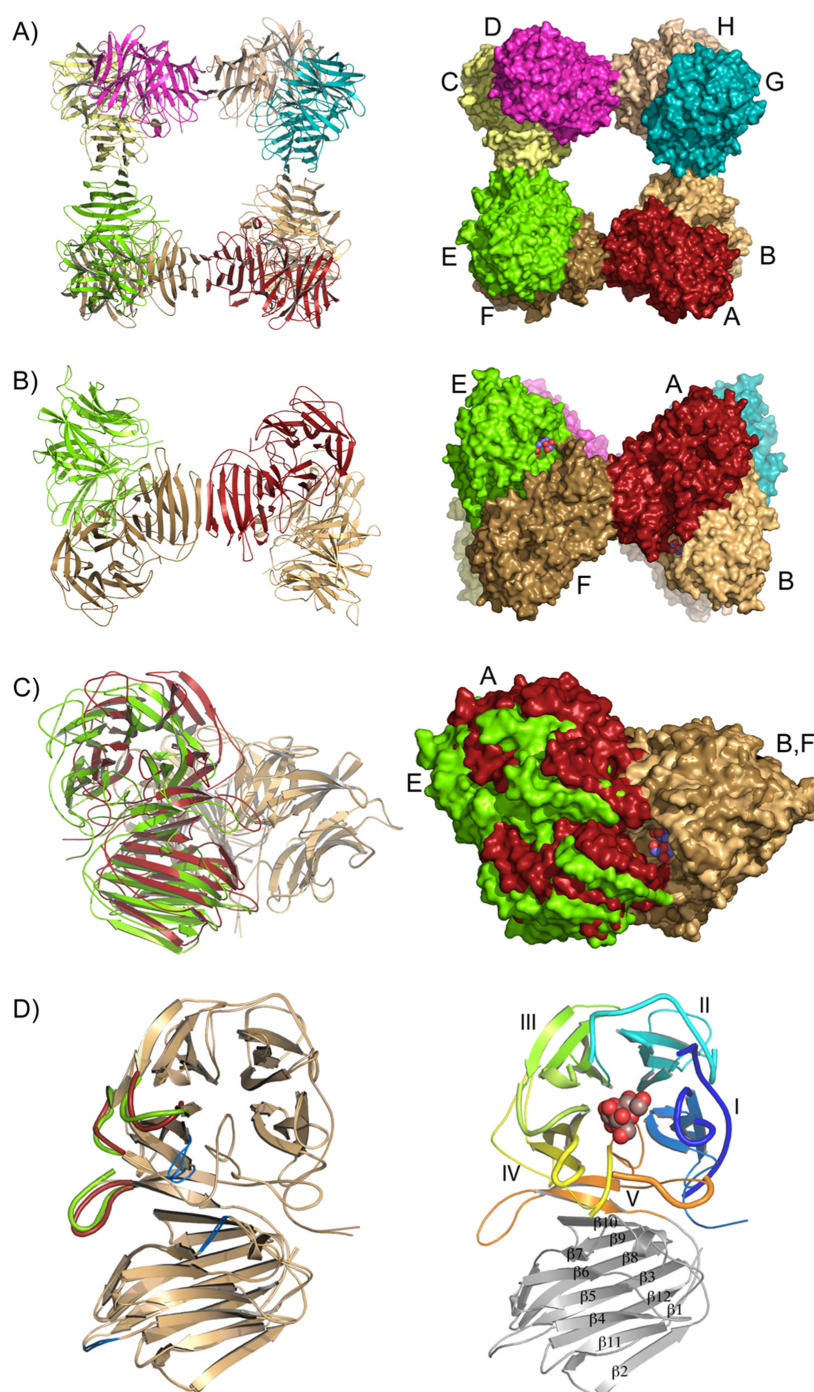


FIGURE 1. Structure of octameric *Slnv*. *a*, view of the *Slnv* octamer in ribbon (*left*) and solvent-accessible surface (*right*) representations, showing each subunit in a different color. In *b*, the octamer is rotated 90°, illustrating that it can be best described as a tetramer of two different kinds of dimers, AB/CD and EF/HG, which are compared by superimposing subunit F on subunit B in *c*. Then, a rotation of 15° would be necessary to bring out subunit E into the A position. Three regions of the polypeptide chain act as hinges (DynDom server), which are colored *blue* in *d* (*left*). The different dimer associations produce local conformational differences at the dimer interface in specific regions represented in *red* (A, B, C, and D) and *green* (E, F, G, and H). *Slnv* folds into two domains, a catalytic β -propeller that is colored according to its five blades (I–V) and a C-terminal β -sandwich domain formed by two antiparallel six-stranded β -sheets (*right*). The putative position of a 1-kestose substrate molecule shown in *c* and *d* is inferred from structural superposition of *Slnv* on the *Cichorium intybus* fructan 1-exohydrolase-1-kestose complex (Protein Data Bank code 2AEZ) to point out the active site cavity.

Structure of *Saccharomyces* Invertase

pare the search model with the program CHAINSAW (28) and a protein sequence alignment of SInv and SoFase. A solution containing eight molecules in the asymmetric unit (molecules A–H) was found using reflections within the 50–3.5 Å resolution range and a Patterson radius of 45 Å, which, after rigid body fitting, led to an *R*-factor of 46%. Crystallographic refinement was performed using the program REFMAC (29) within the CCP4 suite with flat bulk solvent correction and using maximum likelihood target features. Tight local non-crystallographic symmetry (NCS) and jelly body restraints were applied during the first steps of refinement. The free *R*-factor was calculated using a subset of 5% randomly selected structure factor amplitudes that were excluded from automated refinement. Some regions of the polypeptide chain, in particular loops 190–200 and 230–240 and β -strand 228–250, all located at the interface between dimers, were not visible in the electron density of molecules EFGH and were excluded from the model during the first stages of the refinement. Furthermore, two NCS groups were defined composed of molecules ABCD and EFGH, respectively. After iterative refinement and rebuilding of these regions using the program Coot (30), the final $2F_o - F_c$ map showed continuous density for the whole protein. At the later stages, water molecules were included in the model, which, combined with more rounds of restrained refinement, led to a final *R*-factor of 22.9 ($R_{\text{free}} = 23.7$) for all data set up to 3.3 Å resolution. Refinement parameters are reported in Table 1.

Stereochemistry of the models was checked with PROCHECK (31) and MolProbity (32). The figures were generated with PyMOL (33). Analysis of the interfacial surfaces and oligomer stability was done with the Protein Interfaces, Surfaces and Assemblies service (PISA) at the European Bioinformatics Institute (34). Root mean square deviation analysis was carried out with the program SUPERPOSE in the CCP4 package (27).

Activity Assays—Purified SInv was incubated for different times with sucrose, 1-kestose, nystose, raffinose, or inulin (from dahlia tubers) in 100 mM acetate buffer (pH 4.8) at 50 °C. The enzyme was inactivated by heating at 95 °C for 10 min. The calculation of the initial velocity of hydrolysis was based on the kinetics of product release. The products of the reaction were analyzed by anion exchange chromatography using a CarboPac PA100 column (4 × 250 mm) coupled to a pulsed amperometric detector (Dionex) as described previously (23).

RESULTS

The invertase of the yeast *Saccharomyces* was produced in *E. coli*, purified, and crystallized as reported previously (25). We present here the three-dimensional structure of the enzyme solved by molecular replacement at 3.3 Å resolution. The experimental and structure determination details are given under “Experimental Procedures” and in Table 1. The crystals belong to space group P3₁21, with the asymmetric unit containing a complete homo-oligomer of eight subunits. Each chain (A–B–C–D–E–F–G–H) consists of 512 residues with a molecular mass of 58.5 kDa as calculated from its primary structure. The imposition of tight NCS during early refinement led to a model with eight identical subunits. Most of the polypeptide chain exhibited good electron density, but there were some segments that remained undefined, mainly the regions located at the sub-

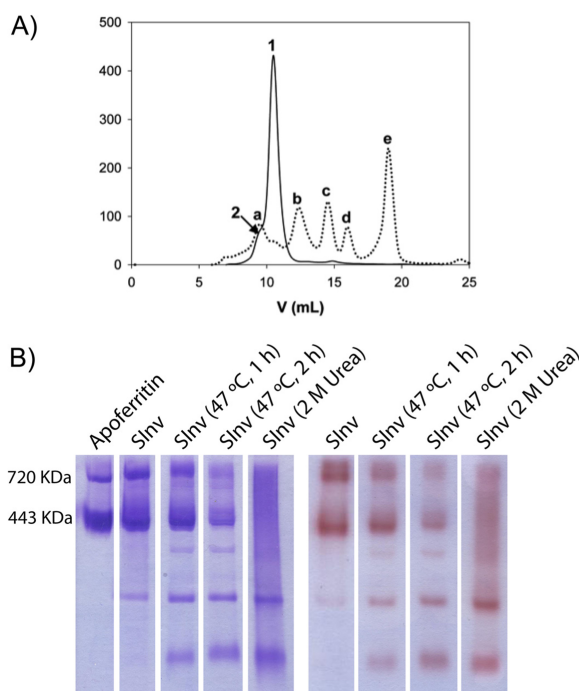


FIGURE 2. Analysis of SInv oligomeric state. *a*, size exclusion analysis of SInv. A sample of purified enzyme (2 mg/ml) in 0.05 M phosphate buffer (pH 7) and 150 mM NaCl was injected onto a Superdex 200 10/300 GL column coupled to an ÄKTA purifier system (GE Healthcare) previously equilibrated with the same buffer. Elution was carried out at a flow rate of 0.5 ml/min for 1.5 column volumes (solid line). SInv eluted mainly as a peak of ~430 kDa (peak 1), indicative of an octameric structure. A shoulder of this peak at lower elution volumes, highlighted with an arrow (peak 2), probably corresponds to higher molecular mass aggregates. Calibration of the column was performed with molecular mass standards (Bio-Rad catalog no. 151-1901) eluted under the same conditions (dotted line): peak *a*, thyroglobulin (670 kDa); peak *b*, γ -globulin (158 kDa); peak *c*, ovalbumin (44 kDa); peak *d*, myoglobin (17 kDa); and peak *e*, vitamin B₁₂ (1.35 kDa). *b*, nondenaturing PAGE analysis of purified SInv performed on 6% polyacrylamide gels. Gels were stained with Coomassie Blue (left), or alternatively, invertase activity was detected by incubating the gels in sucrose and subsequently staining with 1% (w/v) 2,3,5-triphenyltetrazolium chloride in 0.25 M NaOH (right) as described previously (22). Oligomers lower than the octamer appeared after heating SInv at 47 °C or incubating with urea. These treatments also decreased the proportion of the sample that formed higher aggregates. Other external conditions, such as addition of NaCl (0.1, 0.5, and 1 M), dilution (1:3, 1:5, and 1:10), or varying pH (4.5, 7.5, and 8.5), did not alter the pattern of untreated SInv (not shown).

unit interface. Consequently and according to the oligomerization pattern shown by SInv (described below), two strict NCS groups were defined so that chains A–B–C–D and E–F–G–H were refined independently. This scheme led to a continuous electron density in all the subunits, where all the residues could be fitted. The final model showed a root mean square deviation of 1.02 Å between respective C α atoms from both NCS groups, with the differences being restricted to some particular regions of the polypeptide chain, as it will be explained below.

Crystallized SInv Is an Octameric Enzyme—The molecular mass of purified SInv, 428 kDa (see Fig. 2*b*), was consistent with an octamer association as proposed previously (25). Structural analysis showed that it is a flat square-shaped octamer with dimensions of 130 × 130 × 110 Å and is made up of eight subunits related by non-crystallographic 2-fold symmetry par-

TABLE 2**Polar interactions at the SInv interfaces**CD, SInv catalytic domain; β D, SInv β -sandwich domain. Asterisks indicate interactions at the active site pocket.

Molecule B/D		Molecule A/C	
Hydrogen bonds			
CD		CD	
GLN 171 (NE2)	*	GLU 323 (OE2)	
ASN 194 (ND2)		ASP 254 (OD1)	
ASN 255 (ND2)		GLU 195 (OE1)	
SER 257 (OG)		SER 257 (OG)	
ASN 321 (ND2)	*	GLN 171 (O)	
ASN 321 (ND2)	*	TYR 200 (O)	
THR 324 (OG1)	*	GLN 171 (O)	
CD		βD	
GLY 235 (N)	*	SER 439 (OG)	
βD		βD	
ARG 406 (NH1)		GLU 430 (O)	
GLU 410 (N)	*	TYR 433 (OH)	
TYR 433 (OH)		SER 415 (O)	
THR 435 (OG1)		ARG 437 (O)	
ASN 436 (N)		ASN 436 (OD1)	
ARG 437 (NH1)		PHE 434 (O)	
ARG 437 (NH1)		VAL 428 (O)	
PHE 434[O)		ARG 437[NH1]	
Molecule F/H		Molecule E/G	
Hydrogen bonds			
βD		βD	
ARG 406 (NH1)		GLU 430 (O)	
SER 415 (OG)		TYR 433 (OH)	
TYR 433 (OH)	*	SER 415 (O)	
TYR 433 (OH)	*	SER 415 (OG)	
ARG 437 (NE)	*	ASN 436 (OD1)	
ARG 437 (NE)		PHE 434 (O)	
ARG 437 (NH1)		LYS 429 (O)	
ARG 437 (NH2)	*	ASN 436 (OD1)	
Salt bridges			
CD		βD	
ASP 45 (OD2)	*	LYS 385 (NZ)	
Molecule F/G/E/H		Molecule A/B/C/D	
Hydrogen bonds			
βD		βD	
ARG 345 (N)		TYR 358 (OH)	
ARG 345 (NE)		THR 348 (O)	
THR 348 (OG1)		THR 348 (OG1)	
THR 350 (OG1)		TRP 343 (O)	
SER 357 (N)		ASP 361 (O)	
ASN 359 (N)		ASN 359 (O)	
ASP 361 (N)		SER 357 (O)	
SER 363 (N)		ALA 355 (O)	
ALA 355 (O)		SER 363 (N)	
SER 357 (O)		ASP 361 (N)	
ASN 359 (O)		ASN 359 (N)	
ASP 361 (O)		SER 357 (N)	

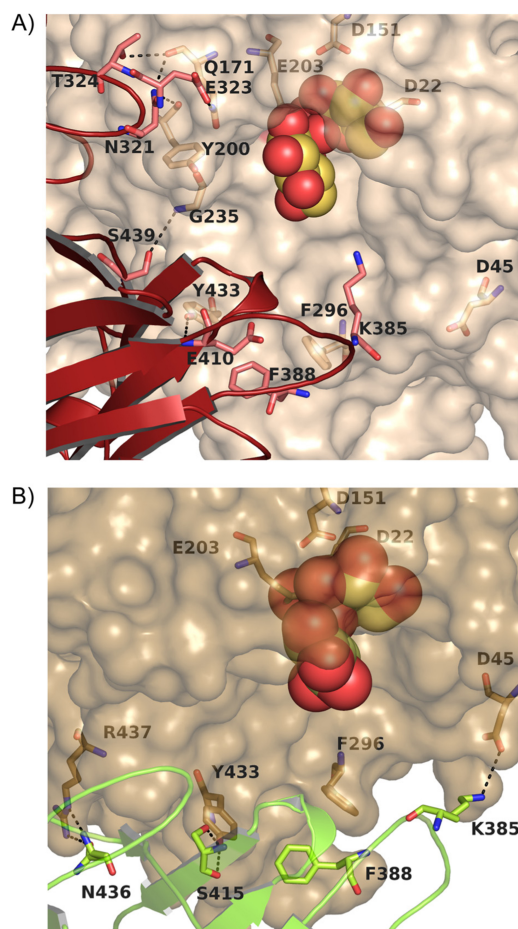


FIGURE 3. Dimer interface at the active site. The octameric SInv active site interfaces are detailed, keeping the same color pattern in Fig. 1c, with one subunit being shown in ribbon representation for clarity. *a*, the AB/CD dimers are tightly made by interactions among both their catalytic and β -sandwich domains. The base of the catalytic pocket is additionally lined by hydrophobic interactions through Phe-388 and Phe-296. *b*, by contrast, the EF/GH dimers interact only through their β -sandwich domains. In addition, the catalytic pocket is also paved by a new salt bridge formed between Asp-45 and Lys-385 from the β -sandwich domain, which lines the cavity. A putative 1-kestose molecule is shown in spherical representation (inferred as explained in the legend to Fig. 1).

allel to the *c* axis in a 222 arrangement (Fig. 1a). Its molecular surface is 136,914 Å², with a total buried surface area of 22,296 Å².

However, this octamer is best described as a tetramer of dimers that oligomerize by intersubunit extension of the two β -sheets that end the β -sandwich domain within each subunit (Fig. 1b). Furthermore, close inspection revealed that SInv forms two classes of dimers, AB/CD and EF/GH, which are located at opposite vertices of the square. The subunits of these two classes of dimers associate differently with each other, and, thus, the EF/GH dimers can be described as an "open" assembly, whereas the AB/CD dimers form a "closed" arrangement. Structural superposition of chain B on chain F shows that a 15° rotation would be necessary to bring monomer A into the E

Structure of *Saccharomyces* Invertase

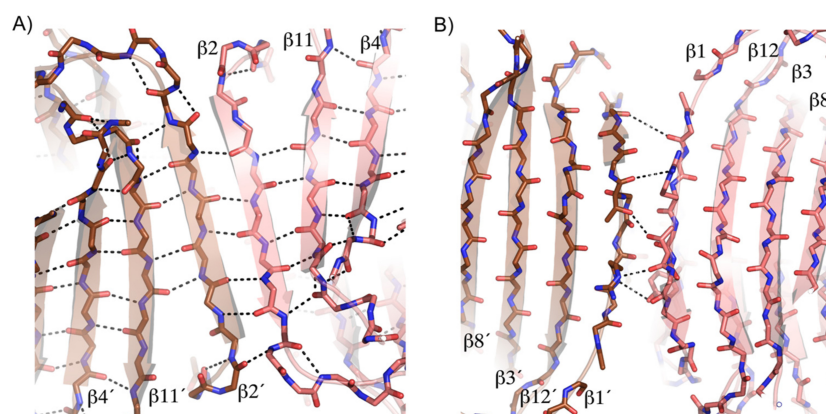


FIGURE 4. **Intermolecular β -sheet.** Both sets of dimers assemble through a similar interface that involves the extension of their corresponding β -sheets between each β -sandwich domain. *a*, view of the internal part at the octamer interface showing that the hydrogen bond pattern centered on $\beta 2$ corresponds to a regular β -sheet. *b*, in contrast, the external wall is defined by polar links between the side chains of the residues at $\beta 1$.

position (Fig. 1c), with three regions of the SInv monomer acting as hinges (colored *blue* in Fig. 1d, *left*).

On the other hand and like the known GH32 members, the SInv subunit folds into two domains, a catalytic β -propeller domain (residues 1–334) and a β -sandwich domain (residues 342–512), linked by a short loop (Fig. 1d, *right*). In turn, the β -propeller domain is assembled from five blades (I–V), each consisting of four antiparallel β -strands (A–D, from the axis toward the outside of the propeller) connected by turns in a classical W pattern. Blade I is the most and blade IV is the least conserved among GH32 family members. The catalytic site is located at the axis of the propeller and is shaped by the loops connecting the different blades (L1–L4) and the turns linking strand B to strand C within each blade (TI–TV). On the other hand, the β -sheet domain has two six-stranded antiparallel β -sheets folded into a β -sandwich topology and is the region presenting the lowest sequence homology among GH32 family members.

The quaternary structure of SInv in solution has been analyzed by different methods. Recombinant SInv subjected to gel filtration chromatography eluted mainly as a peak corresponding to the size of an octamer, although a small fraction of the protein appeared as aggregates of higher molecular mass (Fig. 2a). This tendency to aggregate was also observed on nondenaturing polyacrylamide gels (Fig. 2b). Oligomers of less than eight units (hexamers, tetramers, and dimers) were also detected, and incubation of the enzyme at 47 °C for 1–2 h or treatment with 2 M urea stimulated the dissociation of the octameric form. The enzyme was active in all these different oligomerization states, as shown by the zymogram test (Fig. 2b). Variability in quaternary structure was also observed by ultracentrifuge analysis (data not shown), which indicated that SInv was octameric but showed a decrease in the average molecular mass at the highest centrifugation force (11,000 rpm). These results are in good agreement with those reported for both intracellular and secreted forms of the native enzyme (10, 35).

Different Atomic Interactions Define Dimer Association—Table 2 lists the polar interactions found within each interface of the octamer. First, the AB/CD dimers are tightly associated by

interactions among both their catalytic and β -sandwich domains (Fig. 1c). Fourteen of a total of 32 hydrogen bonds are made between their catalytic domains, mainly through loop L3 and strand D4 and also through a long loop connecting strands C5–D5 (residues 316–328) that makes many interactions near the catalytic pocket, as shown in Fig. 3a. Furthermore, the catalytic pocket of one monomer is surrounded by loops from the β -sandwich domain of the other, and there is a polar interaction from Ser-439 to Gly-235 located at TIV that stabilizes the dimer interface at this region. The base of the catalytic pocket is additionally lined by hydrophobic interactions through Phe-388 and Phe-296 (TV) and direct polar interaction between both β -sandwich domains (Table 2).

By contrast, the EF/GH dimers lack atomic interactions between their catalytic domains, as shown in Table 2, and keep only some hydrogen links between their β -sandwich domains, with half of them not being conserved with respect to the closed dimers. However, the catalytic pocket is also strengthened by a new salt bridge formed between Asp-45 (at loop TI) and Lys-385 (at the loop linking $\beta 3$ – $\beta 4$ of the β -sandwich domain), which encloses a well defined cavity (Fig. 3b). This different pattern of interaction between the subunits of the two kinds of dimers has two direct implications. First, the EF/GH active site has a wider pocket (Fig. 1c). Second, the lack of interactions between the catalytic domains produces a rearrangement in the dimer interface regions that leads to the structural differences of the two kinds of dimers (Fig. 1d).

Both types of dimers assemble through a similar interface that involves the extension of the two β -sheets of each β -sandwich domain (Fig. 4) centered on $\beta 1$ and $\beta 2$, respectively. It is interesting that the sheet constituting the “inner” part of the octamer (Fig. 4a) forms a regular antiparallel intermolecular β -sheet, similar to that reported in multimeric lectins, cytokines, and other proteins (36). These intermolecular interactions between the hydrogen-bonding edges of β -sheets are considered to be a fundamental form of biomolecular recognition (like DNA base pairing) and are involved not only in oligomerization and protein-protein interactions but also in protein aggregation, as they occur in β -amyloid fibril formation (37).

TABLE 3
Analysis of the interfaces

Assembly	Interface area	$\Delta'G$	$\Delta'G$ <i>p</i> value	N_{HB}	N_{SB}	N_{DS}
	\AA^2	<i>kcal/mol</i>				
AB/CD	2664.1	−20.9	0.035	30	0	0
EF/GH	1376.2	−5.5	0.352	17	2	0
CE/AF/DH/BG	767.4	−2.4	0.431	12	0	0

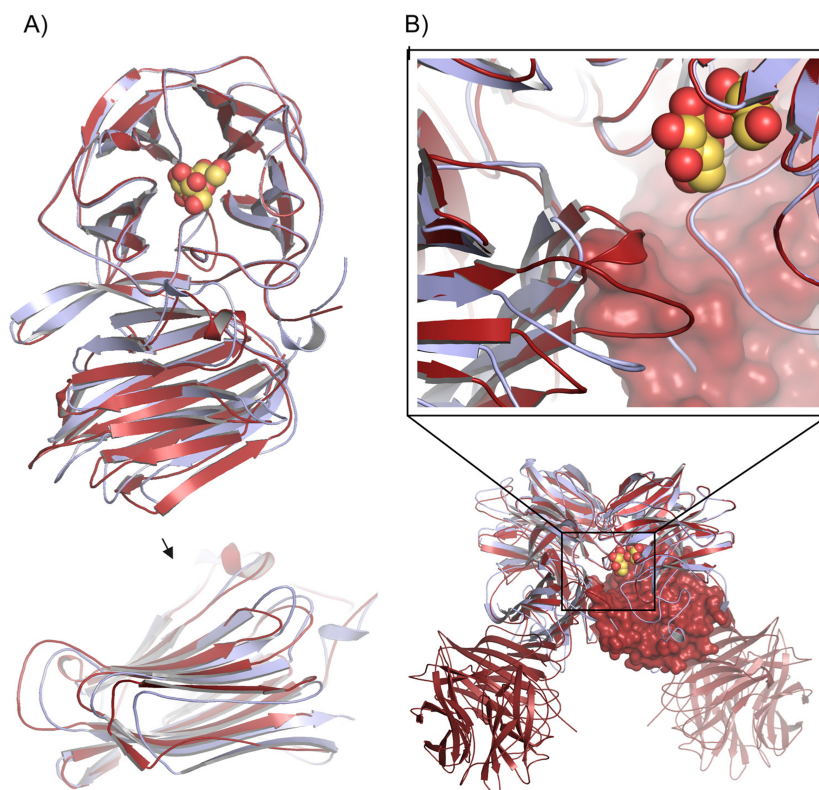


FIGURE 5. Structural differences between Slnv and SoFfase. *a*, structural alignment of the SoFfase subunit (slate) and the Slnv monomer (red) shows well conserved catalytic domains but larger differences in the β -sandwich domain, mostly at strands $\beta 1$ and $\beta 2$, at the base of the β -sandwich domain, which can be attributed to the interactions of the β -sandwich elements in the Slnv octamer. *b*, the β -sandwich of an Slnv subunit is shown in surface representation to highlight the effect of the different arrangement in the active site (magnified) at the dimer interface. The putative 1-kestose position is shown in spherical representation (inferred as explained in the legend to Fig. 1).

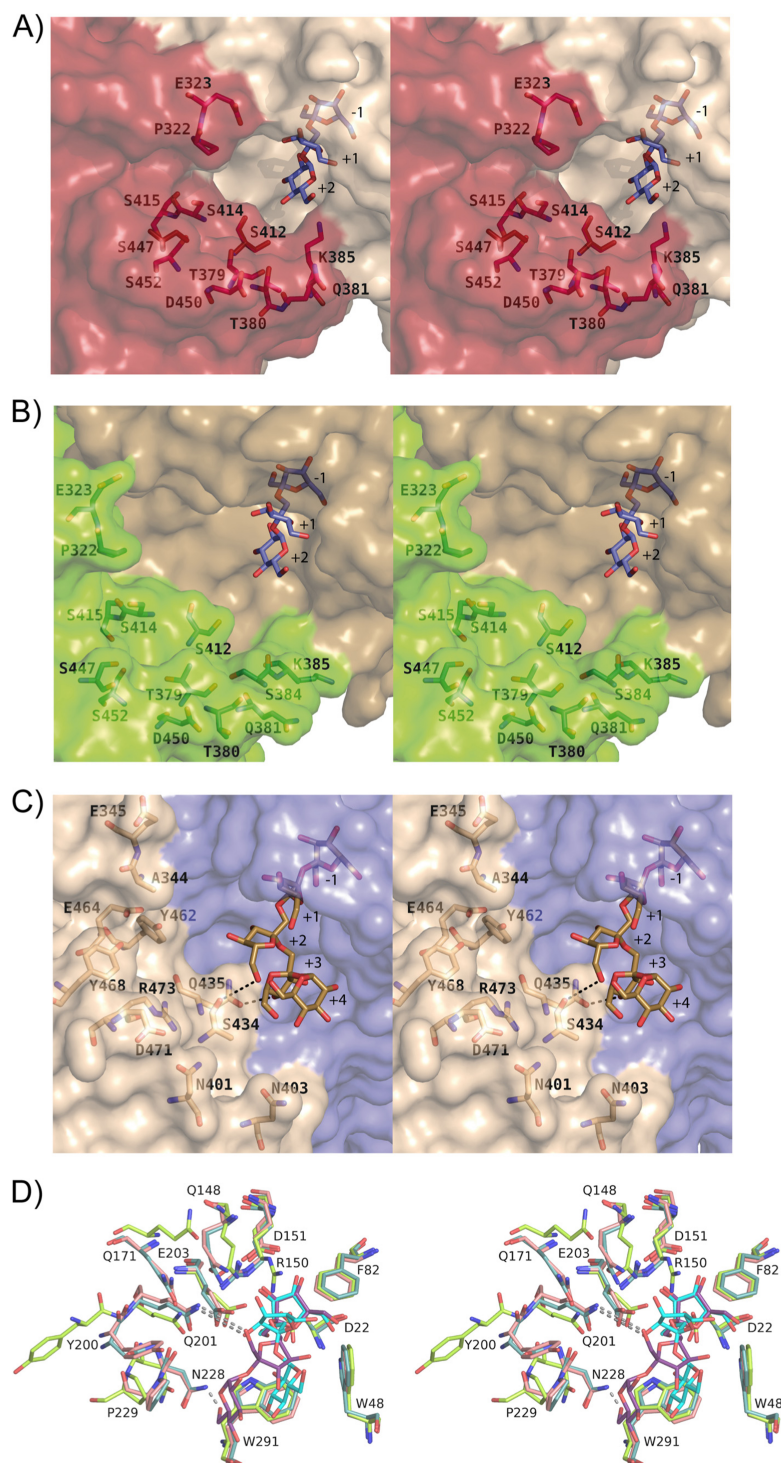
Table 3 summarizes the main features of the different interfaces found in the octamer as analyzed by the PISA program (34). As shown, the interface area of the closed dimers is twice that found in the open dimers, as is the number of polar interactions. The binding energy ($\Delta'G$) is negative in all interfaces, indicating their hydrophobic nature, but the *p* value reveals that the AB/CD interfaces, with lowest $\Delta'G$, have a large hydrophobicity at a higher confidence level. Therefore, only the closed dimers, AB/CD, might be expected to exist in solution. The other interfaces may represent weak interactions (38) existing in higher oligomers of Slnv depending on external conditions. This agrees with the oligomerization behavior of both cytosolic and secreted Slnv isoforms (10, 35).

A Few Changes in Its Sequence Determine the Oligomerization Pattern of Slnv—The specific role of the β -sandwich domain in GH32 enzyme functionality remained elusive for a long time. The first experimental evidence of its implication in

dimerization and substrate binding became available when the first structure from a yeast enzyme, SoFfase, was reported (22). Very recently, the unique role of its β -sandwich domain in substrate recognition has been further demonstrated from the structure of two complexes with long substrates (39). Although SoFfase and Slnv share 68% sequence homology (being 48% identical), SoFfase is a dimeric enzyme, and higher aggregation forms have not been detected.

Structural superimposition of Slnv and SoFfase (Fig. 5) revealed that the catalytic domains are almost identical and that most regions of the β -sandwich domain are very similar. Accordingly, the hydrogen link pattern in the SoFfase dimer is very similar to that in the closed AB/CD Slnv dimers. However, it is remarkable that none of the residues that provide the polar links for the β -sheet extension through $\beta 1$ and $\beta 2$ are conserved in SoFfase. Consequently, this region shows poor structural alignment (Fig. 5). Furthermore, the $\beta 1$ – $\beta 2$ region in

Structure of *Saccharomyces Invertase*



SoFfase has a high content of Lys residues (9/20), which are solvent-exposed at the base of the β -sandwich, yielding a positively charged surface that would prevent dimer association along this region due to electrostatic repulsion effects.

Quaternary Structure Determines Substrate Specificity—The active site of SInv is located at the interface within each pair of dimers. Because of the two different ways subunits can form dimers (Fig. 1c), the active sites in these dimers may have different environments. Thus, the AB/CD dimers form a very narrow pocket of $10 \times 10 \text{ \AA}$ that seems unable to accommodate an oligosaccharide with more than three or four sugar units (Fig. 6a). In contrast, the wider ($20 \times 16 \text{ \AA}$) entrance cavity observed in the EF/GH dimers might allow longer substrates (Fig. 6b), although this could involve a significant, energetically expensive distortion of the polysaccharide.

A comparison of the catalytic pockets of octameric SInv and SoFfase (Fig. 6c) reveals that the shape and size of the SoFfase cavity is more similar to those within the open EF/GH SInv dimers. Nevertheless and more significant, the chemical nature of the residues that form the β -sandwich domain surrounding the active center is quite different in the two enzymes. SoFfase has long chain residues, such as Arg-473, Glu-464, Tyr-462, and Tyr-468 and notably Gln-435, a residue that, together with Ser-412, makes direct polar links with the oligosaccharides at subsite +3 (Ref. 39 and nomenclature according to Ref. 40). Furthermore, Asn-401, Asn-403, and Asp-471 protrude at the entrance of the slot and define the polar boundaries that make accessible the cavity from the solvent. Therefore, most of the relevant SoFfase residues are highly flexible. By contrast, most of the corresponding SInv residues are short chain amino acids, such as Ser-412, Ser-414, Ser-415, Thr-379, Thr-380, and Ser-447. Consequently, the SoFfase active site seems more flexible to accommodate long chain substrates compared with the two catalytic pockets of SInv. This feature may therefore illustrate the structural basis for the activity of SInv as an invertase, whereas SoFfase is in fact an inulinase. As observed when comparing the activity of SInv with different oligosaccharides, the highest efficiency was with sucrose as a substrate, followed by the trisaccharides raffinose and 1-kestose, whereas the tetrasaccharide nystose was hydrolyzed at a much lower rate (Table 4). SInv showed no significant activity with inulin as the substrate.

It has been reported that SInv and SoFfase have some degree of transglycosylating activity, giving 6-kestose as the main product. This activity can be enhanced by mutagenesis (23, 39, 41). In contrast, homologous plant enzymes of the same GH32 family yield mostly 1-kestose. To understand the $\beta(2,6)$ or $\beta(2,1)$ nature of the transglycosylation reaction, the binding site on the putative acceptor sucrose and its orientation relative to the fructose unit in the covalent intermediate must be identi-

TABLE 4
Activity of SInv

Substrate	Activity
	$\mu\text{mol/min/mg}$
Sucrose	520 ± 20^a
1-Kestose	102 ± 11^a
Nystose	36 ± 1^a
Raffinose	187 ± 8^a
Inulin	2.5 ± 0.1^b

^a The substrate concentration was 250 mM.

^b The substrate concentration was 10%.

fied. Fig. 6d illustrates a superimposition of SInv and SoFfase catalytic pockets showing the putative position of the products 1-kestose and 6-kestose. The figure suggests that the nucleophile (Asp-22 in SInv numbering) environment, together with the hydrophobic wall of the pocket (Trp-48, Phe-82, and Trp-291), is coincident in all three cavities. Gln-201 is well positioned to make polar links with the fructose unit at subsite +1 in both transfructosylating products, Asn-228 could link to the glucose moiety of 6-kestose, and the glucose term of 1-kestose would stack with Trp-48. An essential role is thus assigned to Gln-201 to facilitate the transfructosylation reaction by binding the acceptor sucrose, whereas Asn-228 would crucially determine the product specificity as predicted (23) and also as described for SoFfase (39).

In summary, the architecture of the active site, as determined by the way the enzyme monomers are assembled, explains both substrate specificity for hydrolysis (invertase *versus* inulinase activity) and transfructosylation product specificity. Thus, the dimerization mode of SInv modulates its hydrolytic activity, precluding the recognition of long chain substrates. This is more apparent for the closed dimers that would be predominant in aggregation states lower than the octamer or in alternative octameric forms composed exclusively of closed dimers (see "Discussion"). However, the transfructosylating mechanism would be the same in both enzymes.

DISCUSSION

The yeast *Saccharomyces* plays an outstanding role in human civilization as the fermentative agent that produces bread, wine, and beer. *Saccharomyces* owes its predominant position as a fermentative microorganism to a very successful metabolic strategy. In the sugar-rich ecological niches where it dwells, *Saccharomyces* performs a highly efficient mobilization of sugars, which can be channeled to the production of an antiseptic substance (ethanol) that avoids the proliferation of competing microorganisms instead of being used for the production of biomass. The function of a set of enzymes involved in sugar metabolism, including invertase, is critical for the prevalence of *Saccharomyces* in its natural habitat. Despite the important

FIGURE 6. Specificity of SInv. Shown are stereo views of the SInv active site of the AB/CD (a) and EF/GH (b) dimers compared with the SoFfase dimer (c). The putative position of the substrate 1-kestose (inferred as explained in the legend to Fig. 1) is shown in a and b. A fructosyl-nystose molecule (beige) found in the reported E230A SoFfase complex (Protein Data Bank code 3U75) shown in c could enter into the EF/GH active site pocket (b), whereas the narrow entrance at the AB/CD interface (a) would prevent binding of extended fructans. Moreover, in SInv, the short chain residues surrounding the cavity outline a more rigid and therefore less flexible active site to accommodate long and polymeric oligosaccharides. d, structural alignment of SoFfase (slate) in the SInv catalytic pockets (ABCD (red) and EFGH (green)) showing the nucleophile (Asp-22), the intermediate stabilizer (Asp-151), and the acid/base catalyst (Glu-203). The putative positions of the transfructosylation products 1-kestose and 6-kestose are shown in cyan and purple, respectively. The position of 6-kestose has been inferred from superposition of its coordinates extracted from the Cambridge Structural Database (CSD; refcode CELGIC) onto the fructose of docked 1-kestose at subsite -1. The recognition scheme of the donor sucrose moiety giving each product within the AB/CD dimers would be the same as that described for SoFfase, as explained under "Results."

Structure of *Saccharomyces* Invertase

physiological role of SInv and its historical relevance as a model enzyme in early biochemical studies (2), its crystallographic structure had not yet been reported. Although it is secreted by

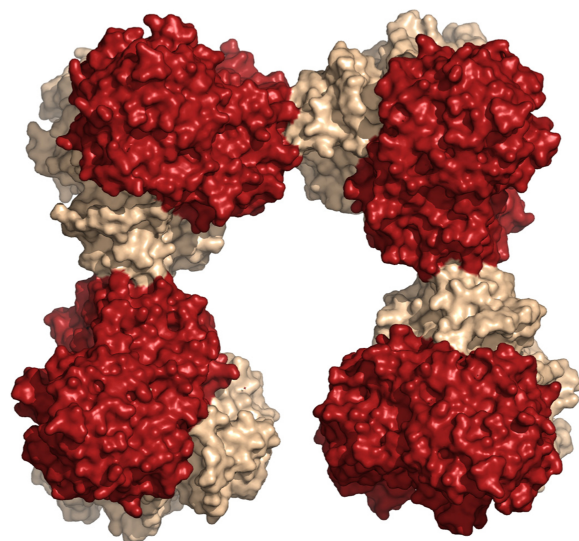


FIGURE 7. **Model of extracellular SInv.** Shown is a putative model of secreted octameric SInv built manually from the AB dimers, with retention of the intermolecular β -sheet. This model reproduces the electron micrographs reported for this isoform (10, 11), which showed rectangles slightly open on one side.

the yeast in large amounts, native SInv appears as a heavy hyperglycosylated protein of heterogeneous molecular mass, which remains trapped in the cell wall (10). It is therefore a rather unsuitable material for crystallization. The biosynthesis and secretion of extracellular proteins by eukaryotes are a complex process in which transit of the nascent polypeptide through the endoplasmic reticulum is coupled with core glycosylation and protein folding. Hyperglycosylation takes place at a later stage in the Golgi and does not affect SInv folding or catalytic activity (42). Due to the fact that core glycosylation is often a necessary condition for proper protein folding, secreted eukaryotic proteins cannot be functionally expressed in *E. coli*. Therefore, the collection of a sufficient amount of protein for crystallization analysis requires either that it is abundantly produced by the organism from which derives or that it is heterologously expressed in a eukaryotic host. Examples of structurally determined SInv homologs of the GH32 family are therefore fungal inulinases abundantly produced by their natural host (17, 20, 22), plant fructosylases heterologously expressed in *Pichia pastoris* (18, 19), and a plant fructosyltransferase abundant in leaves (21). *Saccharomyces* produces an intracellular non-glycosylated version of the invertase (4). The intracellular version, which is synthesized as a soluble protein, is encoded by the same gene as the secreted form (12). Secondary and tertiary structures of both forms are virtually identical, as revealed by CD spectroscopy analysis (35). These considerations led us to carry out heterologous expression of SInv in

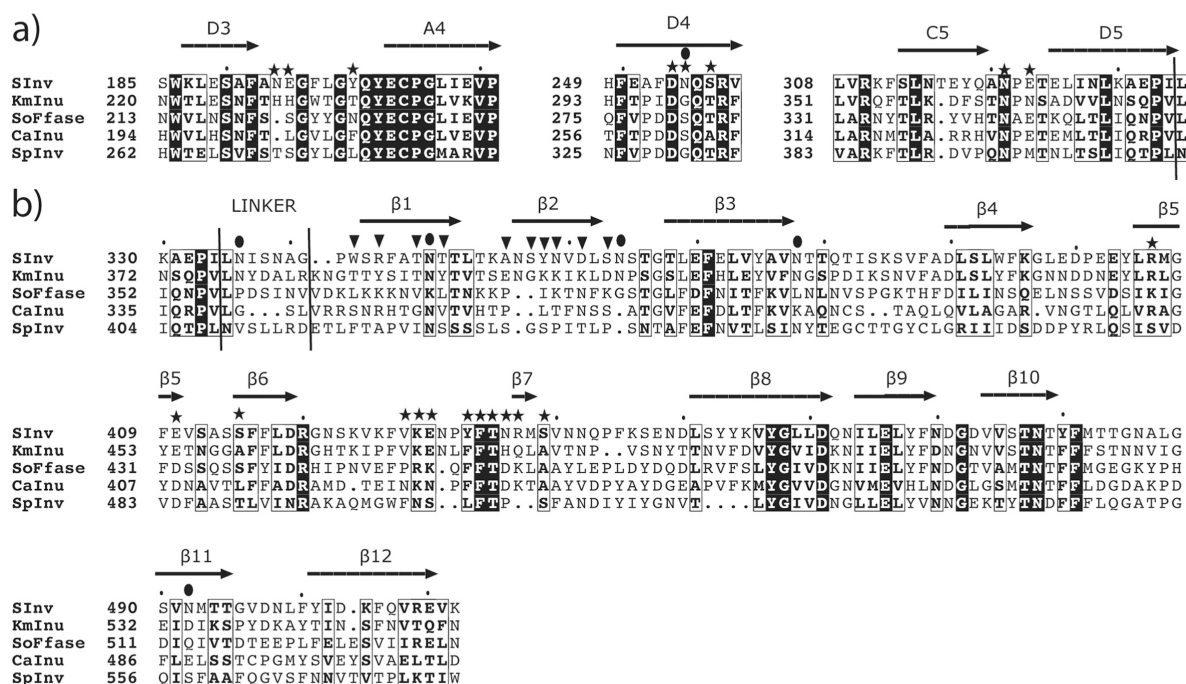


FIGURE 8. **Sequence conservation within GH32 family members from yeast.** The SInv sequence is aligned with other yeast enzymes: *Kluyveromyces marxianus* inulinase (KmInu), *SoFfase*, *Candida sp* KRF1 inulinase (CaInu), and *Schizosaccharomyces pombe* invertase (SpInv). Conserved regions are highlighted (44). The regions of the catalytic domain involved in making the dimer interface are shown in a, whereas the β -sandwich domain is shown in b. Stars specify residues involved in polar links within the AB/CD dimer interface. Inverted triangles indicate residues that make polar links between the β -sandwich domains from contiguous subunits responsible for the formation of the intermolecular β -sheets. Solid circles are potential N-glycosylation sites in extracellular SInv.

E. coli, which yielded suitable protein material for crystallization analysis (25).

Analysis of the crystal structure at 3.3 Å resolution that we have reported here showed that SInv folds into the catalytic β -propeller and β -sandwich domains characteristic of GH32 enzymes. A dimer association that shapes the active site has been found, similarly to that described in the phylogenetically close β -fructosidase from *S. occidentalis* (22). However, in contrast, SInv displays a special assembly of dimers into octamers through extension of each subunit β -sandwich domain. The particular geometry of the octamer generates closed and open dimers that are located alternatively at the vertices of a rectangle. Analysis of the interfaces and binding energy calculation showed that "closed-type" dimers are more stable. Although the octameric form is predominant in a fresh preparation of the enzyme (Fig. 2*a*), the weaker association within the "opened-type" dimer and the dimer-dimer interface may explain the instability manifested by the octamer upon incubation at higher (47 °C) temperatures (Fig. 2*b*) or at a high centrifugal force. Consequentially, recombinant SInv is predominantly an octamer but may exist as a dimer and other oligomeric forms, as has been reported for both the cytosolic and secreted isoforms (10, 11, 35, 43). Furthermore, as seen from the structure reported here, SInv dimerization plays a determinant role in substrate specificity, preventing binding of extended substrates, which explains its invertase character at the molecular level.

Electron micrographs of both internal and secreted SInv (10, 11) show a similar association pattern of spherical units with different oligomerization states: dimers, tetramers, hexamers, and octamers. Interestingly, secreted SInv octamers appear slightly open to one side, whereas intracellular invertase octamers appear mostly as nearly symmetrical closed rectangles, like the structure presented here. The electron micrographs of secreted invertase octamers resemble strikingly the model illustrated in Fig. 7, in which the protein would be composed of closed dimers linked by intermolecular β -sheets.

Although GH32 enzymes are generally highly conserved in the β -propeller and less so in the β -sandwich domain, SInv and SoFase are very similar also in some regions of the β -sandwich domain, namely strands β 8– β 10 and region β 6– β 7. However, strands β 1 and β 2, which build the intermolecular β -sheet within the octamer, are less conserved. In fact, homology alignments of yeast GH32 sequences show that many of them contain a Pro residue at β 1 or β 2 that probably precludes intermolecular β -sheet formation (Fig. 8). In SoFase, an unusually large number of solvent-exposed Lys residues are found in this region, resulting in a positively charged surface that likely interferes with β -sheet dimerization because of electrostatic repulsion effects. Therefore, the association of dimers into octamers seems to be a unique SInv feature.

To conclude, the structure of SInv presented here is an interesting new example of how non-catalytic domains with, unknown function play a role in fine-tuning enzymatic function. The contribution of the β -sandwich domain in building the catalytic pocket of GH32 yeast enzymes has been reported previously. However, a role in higher oligomerization leading to new specificity seems unique to SInv. The production of a complex octameric hyperglycosylated enzyme precludes its diffu-

sion outside the periplasmic space. The secretion of invertase that occurs in many microorganisms represents an evolutionary advantage eliminating a mechanism for sucrose import. The unique quality of *Saccharomyces* is that it keeps the invertase trapped on the cell surface, avoiding its diffusion into the medium, where it would also aid competing organisms. We hope that the resolution of the sophisticated molecular architecture of this enzyme may contribute to the understanding of protein-carbohydrate interactions and to the design of novel, more efficient enzymes for biotechnological purposes.

Acknowledgments—We thank the staff of the European Synchrotron Radiation Facility for providing beam time and for technical assistance at beamline ID23-1. We also thank Simon P. Gough for kindly reviewing the manuscript.

REFERENCES

- Barnett, J. A. (2003) Beginnings of microbiology and biochemistry: the contribution of yeast research. *Microbiology* **149**, 557–567
- Michaelis, L., and Menten, M. L. (1913) The Kinetics of Invertase Action. *Biochem. Z.* **49**, 333–369
- de la Fuente, G., and Sols, A. (1962) Transport of sugars in yeasts. II. Mechanisms of utilization of disaccharides and related glycosides. *Biochim. Biophys. Acta* **56**, 49–62
- Gascón, S., and Lampen, J. O. (1968) Purification of the internal invertase of yeast. *J. Biol. Chem.* **243**, 1567–1572
- Gascón, S., Neumann, N. P., and Lampen, J. O. (1968) Comparative study of the properties of the purified internal and external invertases from yeast. *J. Biol. Chem.* **243**, 1573–1577
- Gilliland, R. B. (1949) A yeast hybrid heterozygotic in four fermentation characters. *Compt. Rend. Trav. Lab. Carlsberg Ser. Physiol.* **24**, 347–356
- Winge, O., and Roberts, C. (1952) The relation between the polymeric genes for maltose, raffinose, and sucrose fermentation in yeasts. *Compt. Rend. Trav. Lab. Carlsberg Ser. Physiol.* **25**, 141–171
- Grossmann, M. K., and Zimmermann, F. K. (1979) The structural genes of internal invertases in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **175**, 223–229
- Carlson, M., Osmond, B. C., and Botstein, D. (1981) *SUC* genes of yeast: a dispersed gene family. *Cold Spring Harb. Symp. Quant. Biol.* **45**, 799–803
- Esmo, P. C., Esmo, B. E., Schauer, I. E., Taylor, A., and Schekman, R. (1987) Structure, assembly and secretion of octameric invertase. *J. Biol. Chem.* **262**, 4387–4394
- Tammi, M., Ballou, L., Taylor, A., and Ballou, C. E. (1987) Effect of glycosylation on yeast oligomer stability. *J. Biol. Chem.* **262**, 4395–4401
- Perlman, D., and Halvorson, H. O. (1981) Distinct repressible mRNAs for cytoplasmic and secreted yeast invertase are encoded by a single gene. *Cell* **25**, 525–536
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238
- Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B., and Czejek, M. (2004) The three-dimensional structure of invertase (β -fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J. Biol. Chem.* **279**, 18903–18910
- Bujacz, A., Jedrzejczak-Krzepkowska, M., Bielecki, S., Redzynia, I., and Bujacz, G. (2011) Crystal structures of the apo form of β -fructofuranosidase from *Bifidobacterium longum* and its complex with fructose. *FEBS J.* **278**, 1728–1744
- Park, J., Kim, M. I., Park, Y. D., Shin, I., Cha, J., Kim, C. H., and Rhee, S. (2012) Structural and functional basis for substrate specificity and catalysis of levan fructotransferase. *J. Biol. Chem.* **287**, 31233–31241
- Nagem, R. A. P., Rojas, A. L., Golubev, A. M., Korneeva, O. S., Eneyskaya,

Structure of *Saccharomyces* Invertase

- E. V., Kulminkaya, A. A., Neustroev, K. N., and Polikarpov, I. (2004) Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. *J. Mol. Biol.* **344**, 471–480
18. Verhaest, M., Van den Ende, W., Roy, K. L., De Ranter, C. J., Laere, A. V., and Rabijns, A. (2005) X-ray diffraction structure of a plant glycosyl hydrolase family 32 protein: fructan 1-exohydrolase IIa of *Cichorium intybus*. *Plant J.* **41**, 400–411
19. Verhaest, M., Lammens, W., Le Roy, K., De Coninck, B., De Ranter, C. J., Van Laere, A., Van den Ende, W., and Rabijns, A. (2006) X-ray diffraction structure of a cell-wall invertase from *Arabidopsis thaliana*. *Acta Crystallogr. D* **62**, 1555–1563
20. Pouyez, J., Mayard, A., Vandamme, A. M., Roussel, G., Perpète, E. A., Wouters, J., Housen, I., and Michaux, C. (2012) First crystal structure of an endo-inulinase, INU2, from *Aspergillus ficuum*: discovery of an extra-pocket in the catalytic domain responsible for its endo-activity. *Biochimie* **94**, 2423–2430
21. Lammens, W., Le Roy, K., Yuan, S., Vergauwen, R., Rabijns, A., Van Laere, A., Strelkov, S. V., and Van den Ende, W. (2012) Crystal structure of 6-SST/6-SFT from *Pachysandra terminalis*, a plant fructan biosynthesizing enzyme in complex with its acceptor substrate 6-kestose. *Plant J.* **70**, 205–219
22. Alvaro-Benito, M., Polo, A., González, B., Fernández-Lobato, M., and Sanz-Aparicio, J. (2010) Structural and kinetic analysis of *Schwanniomyces occidentalis* invertase reveals a new oligomerization pattern and the role of its supplementary domain in substrate binding. *J. Biol. Chem.* **285**, 13930–13941
23. Lafraya, Á., Sanz-Aparicio, J., Polaina, J., and Marín-Navarro, J. (2011) Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. *Appl. Environ. Microbiol.* **77**, 6148–6157
24. Flint, H. J., Scott, K. P., Louis, P., and Duncan, S. H. (2012) The role of the gut microbiota in nutrition and health. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 577–589
25. Sainz-Polo, M. A., Lafraya, A., Polo, A., Marín-Navarro, J., Polaina, J., and Sanz-Aparicio, J. (2012) Crystallization and preliminary x-ray diffraction analysis of the invertase from *Saccharomyces cerevisiae*. *Acta Crystallogr. F Struct. Biol. Crystal. Commun.* **68**, 1538–1541
26. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281
27. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242
28. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643
29. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367
30. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132
31. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
32. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21
33. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA
34. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
35. Williams, R. S., Trumbly, R. J., MacColl, R., Trimble, R. B., and Maley, F. (1985) Comparative properties of amplified external and internal invertase from the yeast *SUC2* gene. *J. Biol. Chem.* **260**, 13334–13341
36. Dou, Y., Baisnée, P. F., Pollastri, G., Pécourt, Y., Nowick, J., and Baldi, P. (2004) ICBS: a database of interactions between protein chains mediated by β -sheet formation. *Bioinformatics* **20**, 2767–2777
37. Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R., and Meier, B. H. (2008) Amyloid fibrils of the HET-S(218–289) prion form a β solenoid with a triangular hydrophobic core. *Science* **319**, 1523–1526
38. Guharoy, M., and Chakrabarti, P. (2007) Secondary structure based analysis and classification of biological interfaces: identification of binding motifs in protein-protein interactions. *Bioinformatics* **23**, 1909–1918
39. Alvaro-Benito, M., Sainz-Polo, M. A., González-Pérez, D., González, B., Plou, F. J., Fernández-Lobato, M., and Sanz-Aparicio, J. (2012) Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β -fructofuranosidase, an enzyme that produces prebiotics. *J. Biol. Chem.* **287**, 19674–19686
40. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557–559
41. Alvaro-Benito, M., de Abreu, M., Portillo, F., Sanz-Aparicio, J., and Fernández-Lobato, M. (2010) New insights into the fructosyltransferase activity of *Schwanniomyces occidentalis* β -fructofuranosidase, emerging from nonconventional codon usage and directed mutation. *Appl. Environ. Microbiol.* **76**, 7491–7499
42. Conde, R., Cueva, R., Pablo, G., Polaina, J., and Larriba, G. (2004) A search for hyperglycosylation signals in yeast glycoproteins. *J. Biol. Chem.* **279**, 43789–43798
43. Reddy, A. V., MacColl, R., and Maley, F. (1990) Effect of oligosaccharides and chloride on the oligomeric structures of external, internal, and deglycosylated invertase. *Biochemistry* **29**, 2482–2487
44. Gouet, P., Courcelle, E., Stuart, D. L., and Métoz, F. (1999) ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308

Structural Analysis of β -Fructofuranosidase from *Xanthophyllomyces dendrorhous* Reveals Unique Features and the Crucial Role of N-Glycosylation in Oligomerization and Activity.

Mercedes Ramírez-Escudero‡, María Gimeno-Pérez‡, Beatriz González, Dolores Linde, Zoran Merdzo, María Fernández-Lobato and Julia Sanz-Aparicio.

The Journal of Biological Chemistry, Vol. 291, No. 13, pp. 6843–6857,

25 de Marzo del 2016

‡ Ambos autores contribuyeron equitativamente a este trabajo.

DOI: 10.1074/jbc.M115.708495

Códigos de acceso del PDB: 5ANN, 5FIX, 5FKB, 5FMD, 5FK7, 5FK8, 5FKC, 5FMB, 5FMC

Interés de la investigación:

La β -fructofuranosidasa de *Xanthophyllomyces dendrorhous* (XdInv) es una enzima altamente glicosilada, cuyas N-glicosilaciones suponen casi el 50% de la masa molecular de la enzima. Además, presenta inserciones importantes respecto a las enzimas descritas, definiendo un subgrupo dentro de la familia GH32. Por tanto, su estudio estructural revelará determinantes estructurales únicos comunes a este subgrupo.

XdInv es capaz de producir FOS con enlaces β (2-6) de tipo neo-FOS, compuestos muy interesantes por sus altas capacidades bifidogénicas. La producción de neo-FOS es la más alta reportada hasta la fecha. Además, esta enzima es capaz de producir hetero-FOS a partir de otros aceptores distintos de la sacarosa. Por ello, la resolución de su estructura y su caracterización molecular es vital para elucidar el mecanismo enzimático, así como para diseñar variantes que puedan producir nuevos FOS más eficientes.

Los objetivos de este trabajo fueron:

- Cristalizar la β -fructofuranosidasa de *Xanthophyllomyces dendrorhous* nativa.
- Establecer las condiciones óptimas de cristalización y difracción del mutante inactivo de la β -fructofuranosidasa de *Xanthophyllomyces dendrorhous* (XdInvD80A) en complejo con sacarosa, 1-kestosa, nistosa, neokestosa, neoerlosa, rafinosa, Fru-HEPES y Fru-BTpro.
- Resolver la estructura tridimensional de XdInv, así como de XdInvD80A y sus complejos.
- Interpretar los datos estructurales y bioquímicos obtenidos de los complejos, e investigar la funcionalidad de los distintos bucles del centro activo en el mecanismo de acción de XdInv.

- Analizar el papel de la extensión del dominio C-terminal y del loop T_{IV} en la actividad y en la especificidad de sustrato.
- Estudiar los determinantes moleculares que generan el dímero de XdInv.
- Examinar el papel que tiene el β-sándwich en la oligomerización y en la especificidad de la actividad de XdInv.
- Analizar el papel del dímero en la degradación de polisacáridos ramificados mixtos.

Las conclusiones de esta investigación son:

- El monómero de XdInv presenta un plegamiento similar al resto de enzimas GH32 resueltas. Sin embargo, contiene dos segmentos extendidos en los extremos N- y C-terminal, con importantes implicaciones estructurales y funcionales. Además, presenta un mayor grado de glicosilación, con 17 puntos de N-glicosilación.
- XdInv también es una enzima dimérica, como las otras dos enzimas GH32 de levadura, resueltas en nuestro grupo. Sin embargo, el dímero presenta una topología única que sitúa ambos centros activos en la misma cara y conectados por una hendidura.
- Esta disposición de los bolsillos catalíticos permitiría la degradación de fructanos mixtos ramificados como el graminano.
- El dímero se forma a través de la N-glicosilación presente en la Asn107 y de la extensión C-terminal del dominio β-sándwich, que se pliega sobre el dominio catalítico y está implicada tanto en la oligomerización como en la actividad.
- El bolsillo catalítico está formado por el dominio β-*propeller* y rodeado por el extremo C-terminal del dominio β-sándwich del mismo monómero. El centro activo presenta dos sitios de unión de sustrato, uno polar que une sustratos tipo inulina y otro hidrofóbico que une los sustratos tipo levano.

La autora de esta tesis ha realizado:

- Desglicosilación del mutante inactivo XdInvD80A.
- Cristalización de XdInvD80A y experimentos de *soaking* de sus complejos.
- Toma de datos de difracción en varias instalaciones de radiación sincrotrón europeas.
- Determinación de la estructura tridimensional de la enzima inactivada en complejo con los distintos ligandos.
- Interpretación de los resultados cristalográficos obtenidos.
- Estudio del papel de la oligomerización en la especificidad de sustrato de XdInv.
- Definición de los determinantes moleculares implicados en cada sitio de unión del centro activo de XdInv.

Structural Analysis of β -Fructofuranosidase from *Xanthophyllomyces dendrorhous* Reveals Unique Features and the Crucial Role of *N*-Glycosylation in Oligomerization and Activity*

Received for publication, December 11, 2015, and in revised form, January 28, 2016. Published, JBC Papers in Press, January 28, 2016, DOI 10.1074/jbc.M115.708495

Mercedes Ramírez-Escudero^{†1}, María Gimeno-Pérez^{§1,2}, Beatriz González[‡], Dolores Linde^{§2}, Zoran Merdzo[§],
María Fernández-Lobato^{§3}, and Julia Sanz-Aparicio^{‡4}

From the [†]Department of Crystallography and Structural Biology, Institute of Physical-Chemistry "Rocasolano," Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid and the [§]Center of Molecular Biology "Severo Ochoa," Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Xanthophyllomyces dendrorhous β -fructofuranosidase (XdINV) is a highly glycosylated dimeric enzyme that hydrolyzes sucrose and releases fructose from various fructooligosaccharides (FOS) and fructans. It also catalyzes the synthesis of FOS, prebiotics that stimulate the growth of beneficial bacteria in human gut. In contrast to most fructosylating enzymes, XdINV produces neo-FOS, which makes it an interesting biotechnology target. We present here its three-dimensional structure, which shows the expected bimodular arrangement and also a long extension of its C terminus that together with an *N*-linked glycan mediate the formation of an unusual dimer. The two active sites of the dimer are connected by a long crevice, which might indicate its potential ability to accommodate branched fructans. This arrangement could be representative of a group of GH32 yeast enzymes having the traits observed in XdINV. The inactive D80A mutant was used to obtain complexes with relevant substrates and products, with their crystal structures showing at least four binding subsites at each active site. Moreover, two different positions are observed from subsite +2 depending on the substrate, and thus, a flexible loop (Glu-334–His-343) is essential in binding sucrose and $\beta(2\text{--}1)$ -linked oligosaccharides. Conversely, $\beta(2\text{--}6)$ and neo-type substrates are accommodated mainly by stacking to Trp-105, explaining the production of neokestose and the efficient fructosylating activity of XdINV on α -glucosides. The role of relevant residues has been investigated by mutagenesis and kinetics measurements, and a model for the transfructosylating

reaction has been proposed. The plasticity of its active site makes XdINV a valuable and flexible biocatalyst to produce novel bioconjugates.

Invertases, or β -fructofuranosidases (EC 3.2.1.26), are biotechnologically important enzymes secreted by many fungi showing wide applications in the food and pharmaceutical industries. β -Fructofuranosidase (XdINV)⁵ from the basidiomycota yeast *Xanthophyllomyces dendrorhous* (also *Phaffia rhodozyma*) is a highly glycosylated protein with a content of 60–70% *N*-linked carbohydrates and a homodimeric active form of 320–380 kDa (1, 2). As other β -fructofuranosidases, XdINV catalyzes the release of β -fructose from the nonreducing termini of various β -D-fructofuranoside substrates such as sucrose, 1-kestose, or nystose (1), and it may also catalyze the synthesis of short-chain fructooligosaccharides (FOS), in which one or two fructosyl moieties are linked to the sucrose skeleton. Most fructosylating enzymes, such as levansucrases (3), inulosucrases (4), and the majority of fructofuranosidases (5), are able to form either $\beta(2\text{--}1)$ (¹F-FOS) or $\beta(2\text{--}6)$ (⁶F-FOS) linkages between fructose units. However, the distinctive property of XdINV is its ability to transfer fructosyl moieties to the glucose unit of sucrose, thus forming neokestose and neonystose as the main transglycosylation products, both neo-FOS with a levan-type structure (⁶G-FOS series). FOS are considered as functional food, acting as prebiotics that selectively stimulate growth/activity of lactobacilli and bifidobacteria in the digestive tract and prevent growth of pathogenic microorganisms. They exert a beneficial effect on human health, contributing to the prevention of cardiovascular diseases, colon cancer, or osteoporosis (6). Neo-FOS possess enhanced properties and chemical stability compared with ¹F-FOS series (e.g. 1-kestose or nystose) typically used as prebiotics (7–9). In addition, XdINV is also able to fructosylate a variety of carbohydrates containing glucose, in particular disaccharides (maltose,

* This work was supported in part by Spanish Ministry of Economy and Competitiveness Grants BIO2010-20508-C04-03/04 and BIO2013-48779-C4-3/4, BioStruct-X Program from the European Community's Seventh Framework Programme Grant FP7/2007-2013, and a Fundación Ramón Areces institutional grant (to the Centro de Biología Molecular "Severo Ochoa"). The authors declare that they have no conflicts of interest with the contents of this article.

The atomic coordinates and structure factors (codes 5ANN, 5FIX, 5FKB, 5FMD, 5FK7, 5FK8, 5FKC, 5FMB, and 5FMC) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) FJ539193.2

¹ Both authors contributed equally to this work.

² Supported by a Spanish doctoral fellowship from the Ministry of Education, Culture and Sports.

³ To whom correspondence may be addressed. Tel.: 34-91-196-4492; Fax: 34-91-196-4420; E-mail: mfernandez@cbm.csic.es.

⁴ To whom correspondence may be addressed. Tel.: 34-91-561-9400; Fax: 34-91-564-2431; E-mail: xjulia@iqfr.csic.es.

⁵ The abbreviations used are: XdINV, *X. dendrorhous* invertase; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; FOS, fructooligosaccharides; GlcNAc, *N*-acetylglucosamine; AaEI, *A. awamori* exo-inulinase; AjFT, *A. japonicus* fructosyltransferase; GH, glycosyl hydrolase; ScINV, *S. cerevisiae* invertase; SoFase, *S. occidentalis* β -fructofuranosidase; Endo H, endoglycosidase H.

Structure and Function of Dimeric XdINV

isomaltulose, isomaltose, and trehalose) and higher oligosaccharides (maltotriose, raffinose, and maltotetraose) but not monosaccharides (glucose, fructose, and galactose) (10). Thus, novel heterofructooligosaccharides with potential applications in the functional food and pharmaceutical industries can be obtained with this enzyme, which makes it an interesting biotechnological target. Further structural analysis of XdINV is necessary to fully understand its particular specificity and to improve its biosynthetic potential.

On the basis of its amino acid sequences, XdINV was classified into the glycoside hydrolase (GH) family GH32 (1), a family including other invertases, β -fructofuranosidases, inulinases, and fructosyltransferases (11), which along with proteins in family GH68 constitute clan GH-J. The enzymes within this clan share a five-blade β -propeller N-terminal domain, in which β -sheets are arranged around a central pocket that accommodates the active site (catalytic domain). Three key acidic residues located in the active site and surrounded by conserved sequences in the GH32 family, NDPNG (D acts as a nucleophile), RDP (D acts as a stabilizer of the transient state), and EC (E acts as an acid base catalyst), are implicated in substrate binding and hydrolysis (12). These enzymes can have hydrolytic or synthetic activity, depending on whether water or an oligosaccharide (the acceptor substrate) accepts the fructose released by fructan hydrolysis (the donor substrate). A distinctive trait of GH32 enzymes is the presence of an additional β -sandwich structure fused at the C terminus of the catalytic domain. In the last decade, GH32 enzymes from bacteria (13, 14), fungi (15–17), and plants (18–20) have been the subject of many structural studies that showed that all are monomeric enzymes. Surprisingly, studies on *Schwanniomyces occidentalis* fructofuranosidase (SoFfase) (21) and *Saccharomyces cerevisiae* invertase (ScINV) (22) showed that these enzymes from yeast are unique in that they form dimers mediated by their β -sandwich domain, which in the case of ScINV associate into higher oligomers. The structural studies have delineated the main structural determinants of activity, the loops and turns connecting the different elements of the β -propeller domain, in the surroundings of the active site, defining the specificity unique to each enzyme (21, 23). However, and remarkably, oligomerization of the yeast enzymes regulates their functionality. Thus, the analysis of the structure of a complex of SoFfase with the polymeric inulin showed the direct involvement of the β -sandwich domain of an adjacent subunit in substrate binding (24).

In this study, we report the three-dimensional structure of the yeast XdINV that discloses novel features and an unusual dimer. Complexes with relevant substrates and products have been obtained, which allowed mapping the catalytic pocket. The role of identified leading residues has been investigated by mutagenesis. Our results give the experimental evidence of the direct involvement of glycosylation in dimerization and activity of a GH32 enzyme and suggest the role of a flexible loop surrounding the catalytic pocket in discriminating substrate specificity.

Experimental Procedures

Organisms, Media, Plasmids, and Mutagenesis—*X. dendrorhous* ATCC MYA-131 was grown in MMM medium as

referred previously (1). *Pichia pastoris* GS115 (*his4*[−], Invitrogen) was used as expression host for the different enzyme variants. It was grown in YEPD medium (yeast extract 1%, peptone 2%, and glucose 2%) at 30 °C. MD media (YNB 1.34%, biotin 4×10^{-5} %, and glucose 2%) was used to select transformants. Expression and induction of proteins in *P. pastoris* were analyzed using BMG and BMM media (both are the same as MD but in potassium phosphate, pH 6.0, and glycerol 1% or methanol 0.5% as carbon sources, respectively). Yeast growth was followed spectrophotometrically at a 600-nm wavelength. *Escherichia coli* DH5 α was used for DNA manipulation and amplification using the standard techniques.

The β -fructofuranosidase *Xd-INV* gene from *X. dendrorhous* (GenBankTM accession number FJ539193.2) comprises an open reading frame of 1995 bp corresponding to a 665-amino acid protein. The QDNS-pIB4 construction contained the 1902-bp fragment of *Xd-INV* responsible for the synthesis of the last 634 amino acids of this protein fused to the 267-bp fragment of the *S. cerevisiae* MF α secretion signal sequence (25). This construct was used as a template to obtain all the mutants generated in this work. Site-directed mutagenesis was carried out using specific primers, including substitutions responsible for mutations N58S, D80A, N107S, E303A, E334S/E334N/E334V/E334Q, Q341N, N342S, H343A/H343T, N471S, and Y659STOP, and following the method described previously (21). DNA sequencing was used to verify that only the desired mutations were present in all the obtained constructs.

Protein Expression and Purification—*X. dendrorhous* β -fructofuranosidase was purified using tangential concentration and DEAE-Sephacel chromatography, as described previously (26). Active fractions were concentrated using Microcon YM-10 (Amicon) filters and stored at -70 °C. β -Fructofuranosidase variants from *X. dendrorhous* (wild type and mutants) expressed in *P. pastoris* were purified as described elsewhere (25). ProtoBlue Safe Colloidal Coomassie-stained (National Scientific) denaturing gel electrophoresis (SDS-PAGE; 8% polyacrylamide) of the samples confirmed the purity of the protein fractions. Protein concentration was determined using the Bio-Rad microprotein determination assay and bovine serum albumin as standard.

After purification of XdINV, protein deglycosylation was performed using endoglycosidase H (Endo H; New England Biolabs) as described previously (26). Deglycosylated protein was subjected to a 1-min ultrafiltration using an Amicon Ultra-4 50K device (Millipore) to remove the Endo H present in the sample and was subsequently concentrated to about 8 mg ml^{−1} using a 10-kDa cutoff membrane.

β -Fructofuranosidase Activity and Kinetics Assays— β -Fructofuranosidase hydrolytic activity was determined by the dinitrosalicylic acid method adapted to a 96-well microplate scale as described previously (25). Standard enzymatic assays were performed using 2% sucrose in 100 mM sodium phosphate buffer, pH 5.5, at 60 °C. Inulin (Orafti GR, BENEOL Ibérica, S.L. Barcelona, Spain) and fructans from *Agave tequilana* (kindly supplied by Dr. Georgina Sandoval, CIATEJ, Guadalajara, Mexico) at 1% were also used in this assay. One unit of activity was defined as that catalyzing the formation of 1 μ mol of reducing sugar/min. Hydrolase activity analysis at 55–80 °C was carried

out under the aforementioned conditions. The thermostability refers to the temperature required for 50% enzyme inactivation (enzyme half-life) after heating about 0.5 units of the pure enzyme at different temperatures (60–85 °C) during 10–120 min and was determined by removing samples at regular intervals and estimating the residual β -fructofuranosidase activity.

For kinetic analysis, the initial velocity was measured in triplicate with 45–30 $\mu\text{g ml}^{-1}$ of enzyme and using 0–80 mM sucrose, 0–60 mM raffinose, or 0–60 mM nystose. The reaction time was 20 min. The plotting and analysis of the curves were carried out using the kinetic module of SigmaPlot software (version 12). Kinetic parameters were calculated by fitting the initial rate values to the Michaelis-Menten equation.

The activity of several XdINV mutants was analyzed by zymogram analysis using non-denaturing gels (6% polyacrylamide) stained with 2,3,5-triphenyltetrazolium-chloride, as described previously (1). Bovine serum albumin was used as reference and negative control.

Crystallization and Data Collection—Crystals of XdINV were grown as described before (26). For data collection, native crystals were transferred to cryoprotectant solutions consisting of mother liquor plus 10% (v/v) glycerol before being cooled in liquid nitrogen. Diffraction data were collected using synchrotron radiation at the European Synchrotron Radiation Facility (ESRF, Grenoble) on ID23-1 beamline. Diffraction images were processed with iMOSFLM (27) and merged using the CCP4 package (28). Crystals of XdINV were indexed in the $P2_12_12$ space group with two molecules in the asymmetric unit and 70% solvent content within the unit cell (Table 1).

The inactive D80A mutant was deglycosylated as reported for native enzyme and tested for crystallization. The best diamond-shaped crystals were grown by mixing equal amounts of 8 mg ml^{-1} protein (in 20 mM Tris-HCl, pH 7) and a solution consisting of 1.3 M sodium citrate and 0.1 M Hepes or 0.1 M BisTris propane, pH 7.5. D80A complexes with 1-kestose or nystose (both from TCI-Europe), raffinose (Sigma), neokestose or neoerlose, both purified from reactions including XdINV as described in references (1, 10), were obtained by soaking into solutions containing 5–30 mM ligand for 5–30 min, whereas the complex with sucrose was obtained by soaking into a 25% (w/v) sucrose solution. Finally, the ternary complexes of D80A-Fru with the corresponding buffer, Hepes or BisTris propane, were obtained in crystals pre-incubated with 30 mM fructose for 5–10 min and then adding different acceptors as maltotriose, isomaltose, or isomaltulose (all from Sigma) that were not found in the crystal. For data collection, all D80A crystals (except the D80A-sucrose complex) were transferred to cryoprotectant solutions consisting of mother liquor plus 10% (v/v) ethylene glycol. Diffraction data were collected using synchrotron radiation on different sources, and images were processed with XDS (29) and merged using AIMLESS from the CCP4 package (28). The data collection statistics are given in Table 1.

Structure Solution and Refinement—The structure of XdINV was solved by molecular replacement using the MOLREP program (30). The best model was the invertase from *Aspergillus japonicus* (PDB code 3LF7), having 34% identity (48% similarity), from which a template was prepared using the program

Chainsaw (31) and a protein sequence alignment of XdINV onto the template. A solution containing two molecules in the asymmetric unit was found using reflections up to 2.5 Å resolution range and a Patterson radius of 40 Å, which after rigid body fitting led to an R -factor of 49%. Crystallographic refinement was performed using the program REFMAC (32) within the CCP4 suite with flat bulk-solvent correction, maximum likelihood target features, and local non-crystallographic symmetry. Free R -factor was calculated using a subset of 5% randomly selected structure-factor amplitudes that were excluded from automated refinement. Extensive model building using the program COOT (33) was combined with several rounds of refinement leading to a model showing a continuous density for the whole polypeptide chain. At the later stages, carbohydrate and water molecules were included in the model, combined with more rounds of restrained refinement that led to a final R -factor of 16.9 ($R_{\text{free}} = 18.8$) for all data set up to 2.14 Å resolution. Refinement parameters are reported in Table 1.

The structure of D80A mutant was solved by molecular replacement using MOLREP and the coordinates of native XdINV as the search model. A first cycle of refinement was performed using REFMAC, combined with model building with COOT. The structures of the D80A complexes were solved by difference Fourier synthesis using these refined coordinates. The ligands were manually built into the electron density maps and were refined similarly, to reach the R -factors listed in Table 1. For neokestose and neoerlose, not present in the Protein Data Bank, a model was built by the on-line biomolecule building program GLYCAM (34). Then PRODRG (35) was used to automatically generate molecular topologies suitable for REFMAC refinement.

Stereochemistry of the models was checked with PROCHECK (36). The figures were generated with PyMOL (37). Root mean square deviation analysis was made using the program SUPERPOSE within the CCP4 package (28).

Polysaccharide Chain Model Building—A model of a $\beta(2\text{--}6)$ -linked fructan chain was constructed using the program GLYCAM (34) and exported in its lowest energy state. This chain was manually docked into the molecular surface of the XdINV dimer. Models containing $\beta(2\text{--}6)$ fructose units, connected with both *gt* or *gg* conformations, were found to fit properly into the cleft connecting the two binding sites.

Results

XdINV Folding Displays Novel Structural Features—We have purified and crystallized the deglycosylated form of the wild-type β -fructofuranosidase XdINV, as reported previously (26). Treatment of this highly glycosylated protein (monomer of 160–200 kDa) with Endo H led to a molecular mass of 66 kDa and reduced drastically the heterogeneity of the sample, allowing us to obtain high quality crystals. We present here its three-dimensional structure solved by molecular replacement at 2.14 Å resolution. Experimental and structural determination details are given under “Experimental Procedures” and in Table 1. It should be noted that the deglycosylated enzyme retains activity, and heterogeneity was reduced without affecting protein quaternary structure, as shown below. The crystals belong

Structure and Function of Dimeric XdINV

TABLE 1

Crystallographic statistics (Values in parentheses are for the high resolution shell)

Crystal data	XdINV	XdINV-D80A / sucrose	XdINV-D80A / 1-kestose	XdINV-D80A / nystose	XdINV-D80A / neokestose	XdINV-D80A / neoerlose	XdINV-D80A / raffinose	XdINV-D80A / Fru/Hepes	XdINV-D80A / Fru/BTpro
Space group	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2
Unit cell parameters									
a (Å)	74.62	74.73	74.53	74.55	74.75	74.57	74.62	74.57	74.43
b (Å)	205.18	205.22	205.019	205.41	205.27	205.07	204.77	205.07	204.74
c (Å)	145.77	146.07	145.99	146.36	145.90	146.59	146.50	146.26	145.560
Data collection									
Beamline	ID23-1 (ESRF)	XALOC (ALBA)	P13 (DESY)	XALOC (ALBA)	XALOC (ALBA)	XALOC (ALBA)	P13 (DESY)	P13 (DESY)	ID23-1 (ESRF)
Temperature (K)	100	100	100	100	100	100	100	100	100
Wavelength (Å)	1.0395	1.10704	0.9786	0.97949	1.10704	1.10704	0.9786	0.9786	0.96863
Resolution (Å)	61.917-2.14 (2.26-2.14)	119.00-2.01 (2.04-2.01)	118.92-1.78 (1.81-1.78)	119.20-1.78 (1.81-1.78)	118.92-2.05 (2.09-2.05)	119.25-1.88 (1.91-1.88)	119.15-1.82 (1.85-1.82)	119.07-1.91 (1.94-1.91)	118.63-1.84 (1.87-1.84)
Data processing									
Total reflections	841615 (103381)	1002740 (49847)	1442954 (71281)	1446083 (70776)	947570 (46665)	1217429 (60761)	1360757 (67001)	1184411 (59459)	864641 (43395)
Unique reflections	117631 (8403)	148093 (7172)	213888 (10480)	215107 (10551)	141191 (6901)	181368 (8861)	200149 (9790)	174276 (8497)	192642 (9476)
Multiplicity	7.1 (6.8)	6.8 (6.8)	6.7 (6.8)	6.7 (6.7)	6.7 (6.8)	6.7 (6.9)	6.8 (6.8)	6.8 (7.0)	4.5 (4.6)
Completeness (%)	99.9 (99.9)	99.0 (98.1)	99.9 (99.8)	100.0 (100.0)	99.9 (99.8)	99.3 (98.6)	99.6 (99.5)	100.0 (100.0)	99.9 (99.9)
Mean <i>I</i> /σ (<i>I</i>)	12.8 (4.3)	9.6 (3.4)	19.5 (3.3)	16.2 (3.6)	10.1 (3.4)	7.4 (3.0)	9.6 (2.6)	10.1 (2.1)	7.4 (1.9)
<i>R</i> _{merge} [†] (%)	10.8 (38.9)	11.8 (55.0)	6.2 (53.4)	6.9 (55.2)	11.5 (59.8)	15.8 (59.1)	6.8 (58.8)	7.2 (48.0)	6.4 (34.3)
<i>R</i> _{pim} ^{††} (%)	3.3 (16.8)	4.9 (22.4)	2.6 (22.1)	2.9 (23.0)	4.8 (24.8)	6.6 (24.2)	2.8 (24.0)	3.0 (19.4)	3.2 (17.4)
Molecules per ASU	2	2	2	2	2	2	2	2	2
Refinement									
<i>R</i> _{work} / <i>R</i> _{free} ^{†††} (%)	16.9/18.8	16.6/18.5	17.0/18.4	16.7/17.8	16.2/18.3	16.8/18.3	16.9/18.5	17.0/18.5	16.3/17.9
N° of atoms/average B (Å²)									
Protein	9616 / 18.6	9610 / 29.82	9610 / 24.75	9610 / 24.08	9610 / 30.86	9610 / 25.78	9610 / 24.07	9610 / 25.55	9610 / 23.34
Carbohydrate	820 / 38.53	913 / 53.94	889 / 45.75	911 / 45.90	842 / 54.79	889 / 47.61	878 / 44.13	875 / 48.033	897 / 46.64
Water molecules	1217 / 27.4	1514 / 41.42	1642 / 36.71	1445 / 36.50	1316 / 41.50	1623 / 39.18	1597 / 35.50	1491 / 35.68	1496 / 34.56
All atoms	11653 / 20.93	12037 / 33.11	12141 / 27.90	11966 / 27.24	11768 / 33.765	12122 / 29.18	12085 / 27.04	11976 / 28.45	12003 / 26.48
Ramachandran plot (%)									
Favoured	95.43	96.00	96.00	96.00	96.00	96.00	96.00	96.00	96.00
Outliers	0	0	0	0	0	0	0	0	0
RMS deviations									
Bonds (Å)	0.009	0.008	0.007	0.007	0.008	0.007	0.007	0.007	0.007
Angles (°)	1.372	1.506	1.402	1.371	1.494	1.405	1.402	1.450	1.413
PDB accession codes	5ANN	5FIX	5FKB	5FMD	5FK7	5FK8	5FKC	5FMB	5FMC

[†]*R*_{merge} = $\sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$, where *I*_i(*hkl*) is the *i*th measurement of reflection *hkl* and [*I*(*hkl*)] is the weighted mean of all measurements.

^{††}*R*_{pim} = $\sum_{hkl} (1/(N-1))^{1/2} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$, where *N* is the redundancy for the *hkl* reflection.

^{†††}*R*_{work}/*R*_{free} = $\sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$, where *F*_c is the calculated and *F*_o is the observed structure factor amplitude of reflection *hkl* for the working/free (5%) set, respectively.

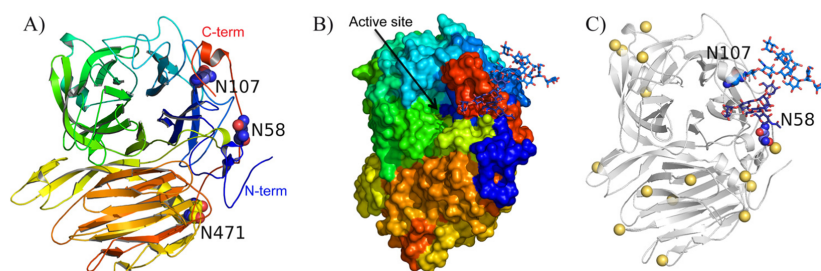


FIGURE 1. Structure of XdINV. A, ribbon diagram of XdINV monomer showing the bimodular molecular architecture, colored in rainbow sequence-code, which folds into a catalytic β -propeller domain and a β -sandwich domain. Two well-ordered glycan chains are observed at positions Asn-58 and Asn-107, whereas Asn-471 is the only position conserved within GH32 from fungi. B, molecular surface of XdINV with the same color code, and the N-glycan chains shown as sticks. C, XdINV is a highly glycosylated protein containing 18 potential glycosylation sites; apart from the two long N-glycans, GlcNAc units have been found in the crystal N-attached to other Asn residues, their location being indicated by small yellow spheres.

to the space group $P2_12_12_1$, with two molecules per asymmetric unit that are related by a non-crystallographic 2-fold symmetry axis. No significant structural differences were observed between the two chains (Ser-38–Tyr-665) that present a root mean square deviation of 0.325 Å on 627 matched C α atoms.

In common with the GH32 enzymes, XdINV is bimodular (Fig. 1) and folds into a catalytic β -propeller domain (residues 57–461) and a β -sandwich domain (residues 473–637), linked by a short α -helix (residues 462–472). The β -propeller has five blades, each consisting of four antiparallel β -strands (A, B, C, and D) with the classical “W” pattern around the central axis, enclosing the active site. The loops linking the different blades (L_1 – L_5), and strand B to C (T_1 – T_5) within each blade, are shaping a deep cavity and are the most variable regions within family GH32 conferring specificity to the different enzymes (Fig. 2). The C-terminal β -sandwich domain is organized in two β -sheets with five strands. However, a unique feature of XdINV is a sort of lace made by an extension of the N terminus (residues 38–56) and a long 30-residue extension at the C-terminal chain (residues 635–665), which increases the interaction between both domains (Fig. 1, A and B). Remarkably, the long C-terminal region makes an extended arm that folds onto blade 1 from the catalytic domain, contouring the active site pocket at its end. As will be explained below, this region is essential for activity and oligomerization.

Endo H treatment is known to cleave the oligosaccharide moieties leaving single N-acetylglucosamine (GlcNAc) units, although poorly accessible glycosylation sites may remain partially glycosylated. XdINV is a highly glycosylated protein that contains 18 potential glycosylation sites in its analyzed sequence. From the electron density maps, GlcNAc units were modeled at 17 Asn positions (represented in Fig. 1C). However, a remarkable feature are the two well-ordered glycan chains that were observed at positions Asn-58, placed at the N terminus extension, and Asn-107, located at T_1 loop in blade 1. Interestingly, the long C-terminal arm is sandwiched between both glycan chains, shaping a platform that is crucial for oligomerization, as described below.

XdINV Is a Dimeric Enzyme—Analytical ultracentrifugation of native and deglycosylated XdINV samples showed an average molecular mass of 260 and 130 kDa, respectively (data not shown), which are consistent with the dimeric state observed by crystallography. The dimer is a flat oblate with dimensions

$135 \times 75 \times 45$ Å (Fig. 3), which is made up of two identical subunits related by a crystallographic 2-fold axis. Its molecular surface is 43733 Å², with an interface of 1037 Å². The interaction between subunits is made by blades 1 and 2 from each catalytic domain, through loops L_1 (residues 135–141), T_{II} (residues 178–182), and L_2 (residues 200–210), and also by the C-terminal region (residues 650–662) (see Figs. 2 and 3). Although the buried surface area is moderate (20 residues are more than 50% buried), and only four direct hydrogen bonds are linking both subunits (Table 2), the interface is further reinforced by the glycan chains linked to each Asn-107, which make a net of intermolecular hydrogen bonds, mainly to loops L_1 and L_2 (Table 2) and also to each other (Fig. 3D). These glycans are filling a tunnel at the dimer interface (Fig. 3C), which results in an increment of the contact surface up to 4150 Å² that should contribute to its stabilization. Moreover, the glycans attached to Asn-58 are packed over the C-terminal region extending further the contact interface to 6140 Å² (Fig. 3B). The contribution of the glycans at the interface and the importance of the atomic interactions explain why treatment of XdINV with Endo H does not remove the glycan chains attached to Asn-58 and Asn-107.

N-Linked Glycans Are Crucial for Oligomerization and Activity—To investigate the precise role that glycosylation plays in the activity of XdINV, positions Asn-58 and Asn-107 were mutated to serine, and properties of the modified protein were explored by zymogram analysis and thermal stability assays. Enzymatic activity of these mutants was also tested. Additionally, Asn-471 was included in this analysis taking into account that this position is conserved among structurally known GH32 enzymes from fungi. Fig. 4A and Table 3 show that N58S and N471S mutations result in active dimeric forms of XdINV. In contrast, mutation N107S causes a total loss of activity that is possibly due to a defective protein folding resulting in a distorted variant of the protein that cannot be detected even by native PAGE (Fig. 4A). Moreover, the lack of a band on SDS-PAGE (Fig. 4B) suggests that mutant N107S might have not been recognized by the secretory pathway. This fact points to the conclusion that glycosylation at Asn-107 is essential for dimer formation, and its removal could be deleterious for proper protein folding. Glycosylation and thermostability are closely related properties in a protein. Thus, the effect of the N58S and

Structure and Function of Dimeric XdINV

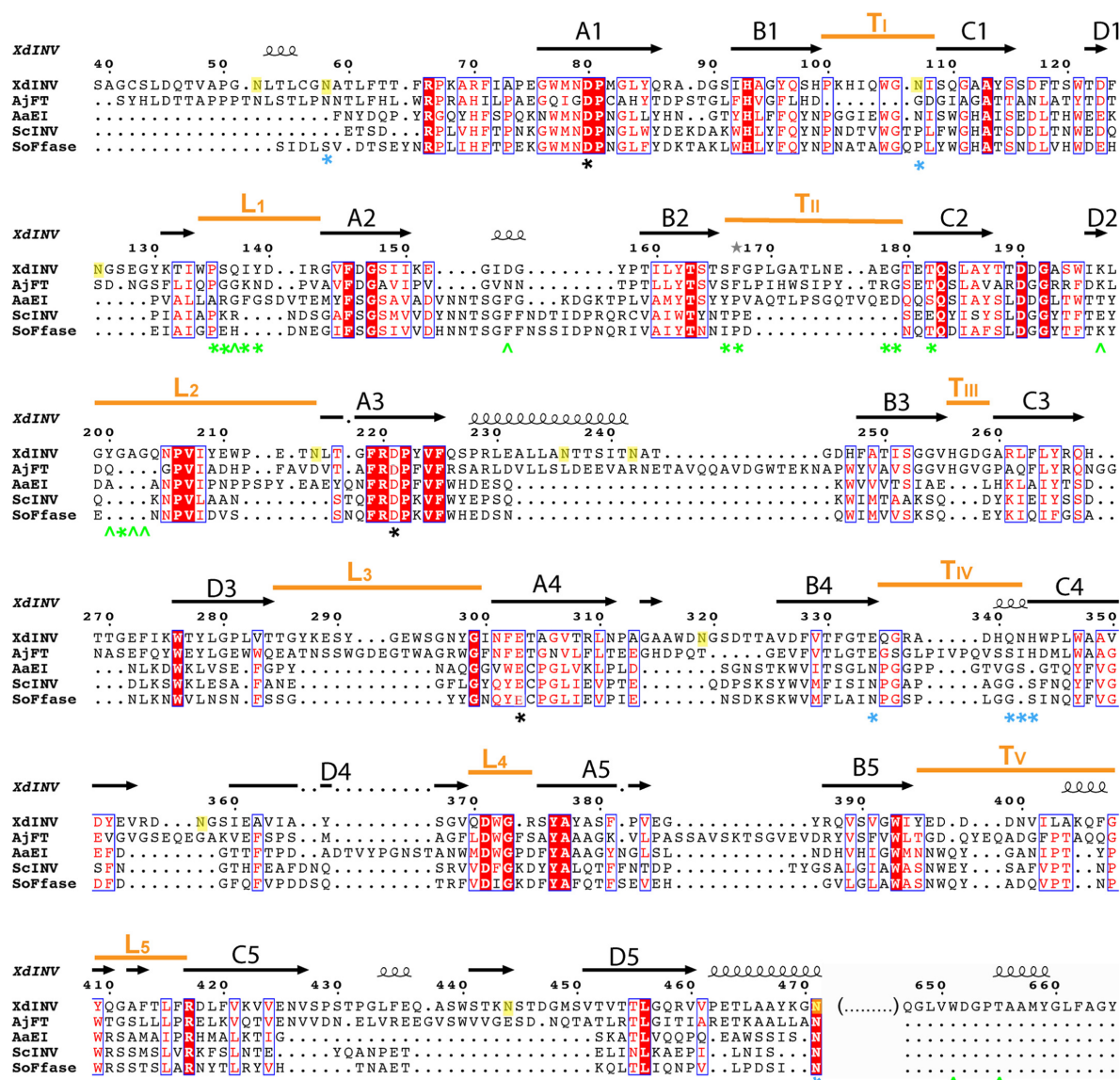


FIGURE 2. Structural alignment of GH32 from fungi. The catalytic domain of XdINV is superimposed to the *A. japonicus* fructosyltransferase, AjFT (PDB code 3LEM), *A. awamori* exo-inulinase, AaEI (PDB code 1Y9G), *S. cerevisiae* invertase, ScINV (PDB code 4EQV), and *S. occidentalis* beta-fructofuranosidase, SoFase (PDB code 3U75), as given by the DALI server (44). The red squares indicate amino acid similarity as calculated by ESPript (45). A–D indicate the beta-strands, from the axis toward the outside of the propeller, within each blade. Loops L1–L5, connecting the different blades of the propeller, and turns T1–Tv, linking strands B to C within each blade, are underlined with an orange bar. XdINV glycosylation sites are highlighted in yellow. Catalytic residues are marked with a black asterisk. Residues of XdINV involved in polar (angle) and hydrophobic (asterisk) interactions within the dimer interface are indicated in green. Residues of XdINV mutated in this work are marked with a cyan asterisk.

N471S substitutions on the thermostability of XdINV was evaluated. Accordingly, these protein variants were purified, and their sensitivities to different temperatures were studied as described under “Experimental Procedures.” The N471S variant behaved similarly to the wild-type protein, displaying maximum activity levels (75–100%) in the range of 60–75 °C, whereas the N58S variant activity decreased to about 40 and 20%, at 70 and 75 °C, respectively (Fig. 4C). In addition, when variants were incubated without substrate in

the range of 60–85 °C for 10–120 min and their activity was assayed, wild-type, N471S, and N58S variants maintained 50% of their activity (half-life) at 71–76, 68–76, and 62–64 °C, respectively (results not shown). The N471S mutation apparently did not affect the thermal stability of the enzyme nor its apparent catalytic efficiency (k_{cat}/K_m) (Table 3), whereas the thermal stability of N58S reduced both thermostability (about 10 °C) and catalytic efficiency with sucrose (4-fold) due to an elevated K_m value. These

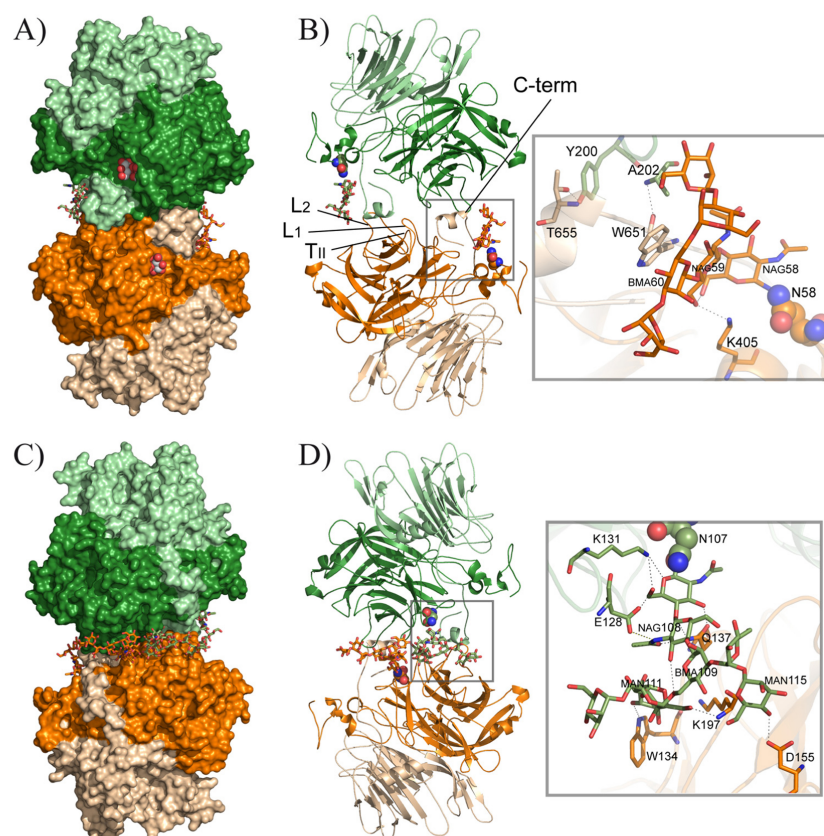


FIGURE 3. **XdINV dimer.** A, molecular surface view of the dimer that is made by association of two subunits through their catalytic domains (dark orange/green) and the C-terminal region of the corresponding β -sandwich domains (light orange/green). A glycerol molecule found in the native protein crystals is shown in sphere representation at each active site pocket. The glycan chain is shown as sticks. B, same view of the dimer in schematic representation, highlighting the regions involved in the interface. Inset, a zoom showing the detailed intermolecular atomic interaction found around the glycan chain attached to Asn-58. C and D the opposite view of A and B showing the dimer and the atomic interactions at the interface around the glycan chain attached to Asn-107. NAG, GlcNAc; MAN, α -mannose; BMA, β -mannose.

TABLE 2
Hydrogen bonds at the dimer interface

Molecule A	Molecule B	Distance \AA
Intermolecular		
Tyr-200 (OH)	Thr-655 (OG1)	2.66
Ala-202 (N)	Trp-651 (O)	3.04
Trp-651 (O)	Ala-202 (N)	2.92
Thr-655 (OG1)	Tyr-200 (OH)	2.68
Trp-134 (NE1)	MAN111 (O3)	2.77
Gln-137 (OE1)	NAG107 (O3)	2.60
Gln-137 (NE2)	NAG107 (O4)	3.00
Gln-137 (NE2)	NAG108 (O7)	3.01
Asp-155 (OD2)	MAN115 (O4)	2.70
Lys-197 (N)	MAN111 (O6)	3.05
Intramolecular		
Glu-128 (OE1)	NAG108 (N2)	2.85
Glu-128 (OE1)	NAG107 (O6)	2.70
Lys-131 (NZ)	NAG107 (O5)	2.74
Lys-131 (NZ)	NAG107 (O6)	2.96
Lys-405 (N)	NAG059 (O6)	2.90
NAG108 (O3)	BMA109 (O6)	2.98
NAG58 (O6)	NAG59 (O6)	3.45
NAG59 (O3)	BMA60 (O6)	3.30

results indicate that glycosylation at Asn-58 contributes to stabilize the structure of XdINV, whereas the role of that attached to Asn-471 is unclear.

XdINV Presents Unique Traits within Known GH32 Enzymes from Fungi—The three-dimensional structure of five other GH32 members from fungi has been reported as follows: the *Aspergillus awamori* exo-inulinase (AaEI (15)), the *A. japonicus* fructosyltransferase (AjFT (16)), the *Aspergillus ficuum* endo-inulinase (17), the *S. occidentalis* β -fructofuranosidase (SoFfase (21)), and the *S. cerevisiae* invertase (ScINV (22)). XdINV shares the highest identity (37%) with AjFT, whereas 20–25% identity is found with respect to the other enzymes, a value equivalent to that observed when XdINV is compared with the GH32 enzymes from bacteria and plants. Structural alignment of the catalytic domain of XdINV onto the fungal exo-enzymes (all but *A. ficuum* endo-inulinase) is shown in Fig. 2. As observed, both XdINV and AjFT present long loops linking the different β -sheets, and remarkably, AjFT presents also the inserted region at the N terminus described in XdINV but not the C-terminal extension. In contrast, Asn-471, at the segment linking the β -structure to its catalytic domain, is the sole conserved Asn residue in all the sequences, although Asn-52 and Asn-58 are present in AjFT and Asn-107 is only conserved in AaEI. Consequently, the glycosylation pattern of XdINV and the concurrence of Asn-107 and the C-terminal extension, both

Structure and Function of Dimeric XdINV

shown to be key determinants in oligomerization, are unique to this protein. Moreover, as can be observed in Fig. 2, loops L₁, T₁₁, and L₂ are highly variable in sequence, and the residues involved in polar links or interatomic interactions at the dimer interface are not conserved.

It is worth noting that from all GH32 enzymes, only the reported structures from yeasts, SoFfase and ScINV, revealed the existence of dimeric forms. However, the dimerization pattern of those enzymes was different from that observed in XdINV (Fig. 5), with blades 4 and 5 involved in the interface of

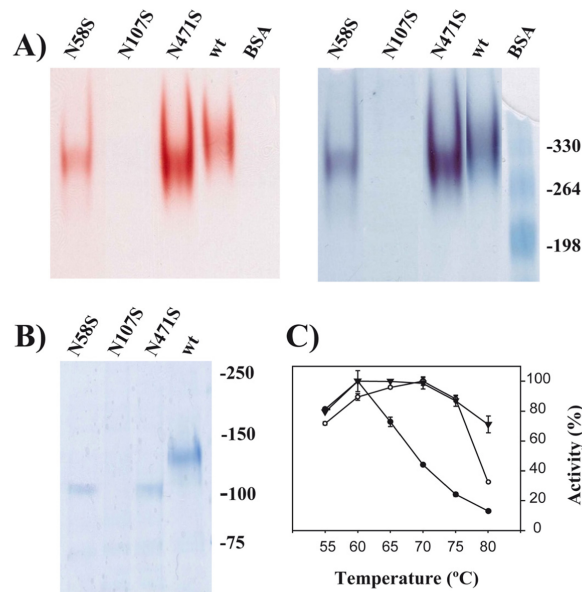


FIGURE 4. Role of glycosylation in XdINV. A, PAGE analysis of the XdINV variants. Culture filtrates (200 ml) of the *P. pastoris* transformants expressing the wild-type (wt) or the indicated XdINV mutants were concentrated through 50000 M_w cutoff (PES membrane) and applied to DEAE-Sephacel chromatography. Active fractions eluting at 0.1 M NaCl were concentrated to 0.2–0.5 mg/ml. The enzyme activities (10 μ l; about 2–5 μ g of proteins) were revealed *in situ* (left) and subsequently stained using colloidal Coomassie (right). The mutant N107S was processed in the same way, but 20 μ l of the final concentrate was analyzed. B, referred proteins were analyzed by SDS-PAGE. Numbers at the right in A and B indicate the positions of the bovine serum albumin (BSA) molecular masses and the weight markers used as control (in kDa), respectively. C, temperature dependence profiles. Activity of the wild-type (filled triangles), N58S (filled circles), and N471S (open circles) protein variants were evaluated at the indicated temperatures. Each value represents the average for four independent measurements. Standard errors are indicated.

TABLE 3

Kinetic analysis of XdINV mutants

The k_{cat} values were calculated assuming a protein molecular mass of 66 kDa (unglycosylated monomer). The \pm sign refers to standard error curve fit using the kinetic module of SigmaPlot version 12. The mutants D80A, N107S, E303A, E334S/E334V/E334N, Y659STOP gave no activity using sucrose. WT, wild type; ND, not detected; blank, not evaluated.

XdINV variant	K_m (mM)			k_{cat} (s ⁻¹)			k_{cat}/K_m (mM ⁻¹ s ⁻¹)		
	Sucrose	Raffinose	Nystose	Sucrose	Raffinose	Nystose	Sucrose	Raffinose	Nystose
WT	1.9 \pm 0.3	7.7 \pm 1.6	21.7 \pm 3.6	911 \pm 47	291 \pm 51	148 \pm 25	479 \pm 68	56 \pm 10	72 \pm 12
N58S	6.1 \pm 1.7			711 \pm 200			117 \pm 33		
E334Q	3.6 \pm 0.6	15.5 \pm 3.4	1.9 \pm 0.4	91 \pm 15	7 \pm 1	46 \pm 10	25 \pm 4	0.4 \pm 0.1	23 \pm 5
Q341N	1.8 \pm 0.4	8.4 \pm 2.0	30.6 \pm 6.9	377 \pm 76	171 \pm 40	71 \pm 16	212 \pm 43	20 \pm 5	2 \pm 1
N342S	7.3 \pm 1.3	2.9 \pm 0.7	12.7 \pm 2.1	1070 \pm 189	165 \pm 41	65 \pm 11	147 \pm 26	58 \pm 14	5 \pm 1
H343A	8.3 \pm 1.5	7.5 \pm 1.5	ND	105 \pm 19	32 \pm 6	ND	13 \pm 2	4 \pm 1	ND
H343T	12.6 \pm 3.3	12.9 \pm 3.5	ND	855 \pm 223	101 \pm 27	ND	68 \pm 18	8 \pm 2	ND
N471S	3.4 \pm 0.9			1151 \pm 307			339 \pm 91		

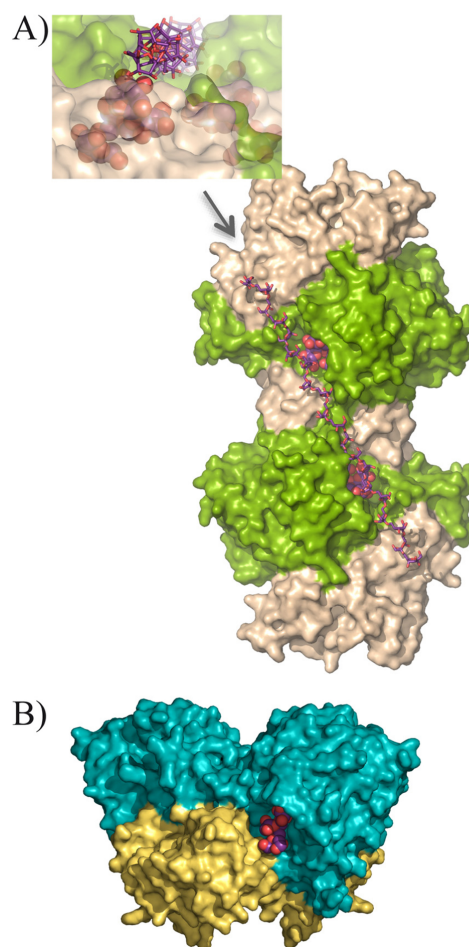


FIGURE 5. XdINV versus SoFfase dimers. A, molecular surface of XdINV dimer showing the two catalytic domains in light green, and the corresponding β -sandwich domains are colored in sand. Each active site cavity is filled with a nystose molecule (in spheres) as found in the complexes obtained in this work and described below. A modeled β (2–6)-linked fructose chain (in sticks) has been manually docked into the long crevice that encompasses the two active sites and beyond. Inset, a detail of this crevice as viewed from the position indicated by the arrow, i.e. along the modeled levant-type fructose chain; a nystose molecule (spheres) is visible at one of the active sites. B, molecular surface of the SoFfase dimer showing the catalytic domains in cyan and their β -sandwich domains in yellow. Each active site is located at one face of the dimer, and only a fructosyl-nystose molecule found in the crystal (24) is visible as spheres.

the SoFase and ScINV dimers through loops L_3 , L_4 , and the segment linking C5-D5 (Fig. 2). Nevertheless, the β -sandwich domain is involved in the assembly of the two types of dimers and also in shaping the active site cavity, although through an evident different pattern. Thus, this feature seems to be a common trend of the GH32 enzymes from yeast, and remarkably, it is the only case in which a role has been assigned to the β -sandwich domain within the GH32 family enzymes. A notable feature of XdINV is that, unlike SoFase, the particular architecture of its dimer houses the active sites at the same face and separated 30 Å away. A close inspection of the molecular surface reveals a long crevice encompassing the two active sites and beyond. Therefore, it is appealing to suggest that this disposition might be conferring the XdINV dimer the additional ability to accommodate a putative polymeric substrate, as illustrated in Fig. 5. In this way, the XdINV cleft could be able to allocate, for instance, a linear $\beta(2\text{--}6)$ -linked fructose chain with $\beta(2\text{--}1)$ branches, characteristic of mixed fructans or graminans (38), its $\beta(2\text{--}1)$ branches being hydrolyzed at each catalytic pocket. In agreement with this, XdINV showed a small but significant hydrolase activity on fructan polymers, showing a moderate activity on inulin (72 units/mg *versus* 960 units/mg on sucrose) as is observed in other GH32 enzymes. Furthermore, XdINV is also able to hydrolyze agave fructans (15 units/mg), which consist of a complex mixture of fructose-based polymers containing principally $\beta(2\text{--}1)$ linkages but also $\beta(2\text{--}6)$ (39). Although this is not evidence that branched substrates can bind to both sites, the channel-like topology of the dimer surface is, indeed, a remarkable trait of XdINV.

Structure of the XdINV Complexes Discloses the Binding Subsites—Inactivated XdINV D80A mutant was crystallized and used for soaking experiments with different donor substrates, including sucrose, 1-kestose, neokestose, and nystose; from these, 1-kestose and neokestose are also transfructosylation products. As has been reported recently (10), XdINV is able to fructosylate glucose-containing oligosaccharides as maltose, which produces neoerlose. To depict this activity, this trisaccharide has been also included in the soaking experiments. Furthermore, in an attempt to capture a putative intermediate complex, mutant crystals were pre-incubated with fructose before being transferred to solutions containing tested acceptors as maltotriose, isomaltulose, isomaltose, and raffinose. However, only raffinose was found at the active site, but it was occupying the catalytic pocket in a donor substrate mode. Interestingly, fructose and a molecule of Hepes or Bistris propane, both from the buffer, were found at the active site in some of the crystals mimicking the location of an acceptor for the fructose in a putative intermediate complex, as will be commented on below.

The catalytic site is buried in a deep pocket 20 Å deep and about 10×10 Å width. All the obtained complexes show clear electron density at the active site, which allowed unambiguous ligand modeling that situated the terminal fructose at the bottom and in contact with the catalytic residues (Fig. 6). This fructose at subsite -1 keeps the same atomic interactions and features previously described for other complexes from GH32 enzymes (16, 19, 24). Remarkably, a net of well ordered water molecules is found in all complexes, contributing to substrate

recognition. The only significant differences found in the protein upon substrate binding are located at loop T_{IV} (residues Glu-334–His-343), which is rather disordered in the free enzyme and is stabilized in the complexes making crucial interactions, as explained below.

The glucose bound at subsite $+1$ in the complex with sucrose, neokestose, and raffinose is tightly fixed by direct hydrogen bonds to the acid/base catalytic Glu-303 (through its O2), and to Glu-334 (O2 and O3) and His-343 (O2), both from loop T_{IV} . The glucose is further stabilized by interaction of its O4 to Leu-170 and Ala-172 from loop T_{II} , through several water molecules. In the case of 1-kestose and nystose, having fructose at this subsite, and neoerlose, with a $\beta(2\text{--}6)$ linked glucose, a water molecule is occupying the position of the glucose O2 observed in the sucrose complex, therefore mediating the same interaction pattern to Glu-334 and His-343.

Subsite $+2$ is made by stacking interaction to Trp-105 from loop T_I , which allocates the corresponding sugar unit in a variety of modes revealing a noticeable plasticity. Apart from this hydrophobic interaction, there are many hydrogen bonds to different residues, always through numerous water molecules and, also, many intramolecular hydrogen bonds fixing each substrate in a precise conformation. First of all, the C-terminal extension is making a polar interaction through Leu-661 main chain, conserved in all the complexes. But additional interactions depend on the type of substrate and, thus, neokestose, raffinose, and neoerlose are accommodated between loop T_{II} and the C-terminal segment by multiple polar interactions to Leu-170, Ala-172, Gly-660, and Leu-661. By contrast, 1-kestose and nystose are oriented toward loop T_{IV} and interacting with Gln-341. Finally, subsite $+3$, occupied only in the nystose complex, is mainly shaped by loop T_{IV} and shows direct polar interactions of the terminal glucose with Gln-335, Asn-342, and His-343.

In summary, subsite $+1$ is common to all complexes, being shaped by T_{II} through Leu-170 and Ala-172 main chain and by T_{IV} through Glu-334 and His-343. Subsite $+2$ is created by stacking to Trp-105 and polar interaction to the main chain of the C-terminal region (Gly-660 and Leu-661), but two different sites are observed depending on the substrate, and thus, inulin-type ligands are accommodated by loop T_{IV} through Gln-341, whereas levan or neo-type ligands are allocated by loop T_{II} (Leu-170 and Ala-172) (Fig. 7). An additional subsite $+3$ for inulin-type substrate is formed essentially by loop T_{IV} (Gln-335, Gln-341, Asn-342, and His-343). In summary, the structure of the XdINV complexes presented here suggests a role of loop T_{IV} in regulating specificity of this protein, especially through residues Glu-334, Gln-341, and His-343.

Mutational Analysis Unveils Key Residues Modulating Activity—The role of the leading residues identified in the complex analysis was investigated by kinetic analysis of mutants that is given in Table 3. First, the expected lack of activity observed in mutants D80A and E303A is consistent with its catalytic role as nucleophile and electron/proton donors, respectively. But, remarkably, the presence of glutamate at position 334 seems to be also required for sucrose hydrolysis, as its replacement by serine, valine, or glutamine is deleterious for activity and leads to fully inactivated enzymes, revealing the

Structure and Function of Dimeric XdINV

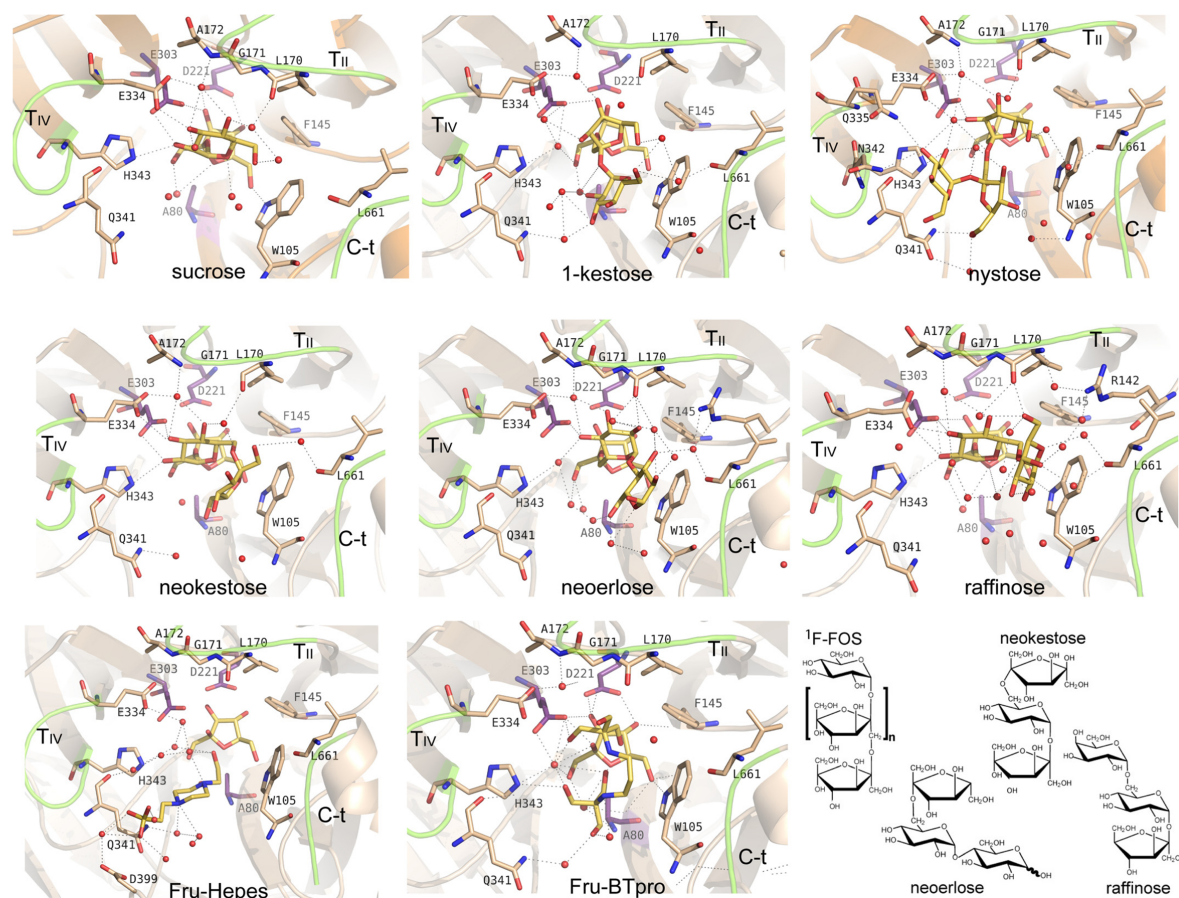


FIGURE 6. Structure of the XdINV complexes. Close-up view of the active site showing the relevant residues making interactions with each ligand. Main polar interactions are represented as *dashed lines*. The interactions at subsite -1 are conserved and only shown in the XdINV-sucrose complex. The catalytic residues are highlighted in *purple* and the referred ligands in *yellow*. Chemical structure of the sugars is given in the scheme, the ¹F-FOS is 1-kestose or nystose, when $n = 1$ or 2; sucrose, $n = 0$.

major role of Glu-334 in binding sucrose. Only the mild E334Q replacement gives detectable activity on sucrose, and interestingly, this mutant keeps a significant activity on nystose. As described above, Glu-334 is making a direct hydrogen bond to the sucrose O2 and O3 atoms, from the glucose located at site +1, and therefore, some effect upon its replacement was expected in the hydrolysis of this substrate. The presence of an acidic residue at this position seems necessary for sucrose binding and hydrolysis in many clan GH-J enzymes, and as such, the equivalent D239A mutant of *Arabidopsis thaliana* invertase exhibited a 6-fold increase in K_m against sucrose (40). Moreover, although this Asp/Ala replacement compromised its ability to degrade sucrose, the enzyme retained fructan hexohydrolase activity (41). Thus, from sequence analysis of plant GH32 enzymes, it was proposed that members having an Asp-239 homolog have sucrose as the preferential donor, whereas enzymes lacking this residue use fructans as their donor substrate (41). In agreement with this role in regulating specificity, kinetic analysis of SoFase mutants at Gln-228 position, structurally related to the *A. thaliana* Asp-239, revealed that Gln-

228 is essentially responsible for conferring affinity for acceptor sucrose in the transfructosylation reaction (24). However, in that case, even the SoFase Q228V and Q228T mutants retained moderate invertase activity. Consequently, the XdINV Glu-334 plays a more essential role in sucrose binding than that reported for the yeast SoFase, resembling the behavior described for plant GH32 enzymes.

Mutants H343A/H343T also showed drastically reduced activity on the assayed substrates. As observed in the complexes, His-343 establishes a direct hydrogen bond to the glucose O2 from sucrose, and in agreement, there is a concomitant decreased binding affinity as revealed by its K_m value. This position is equivalent to Lys-242 from the *A. thaliana* invertase that, together with Asp-239, is proposed to be a hot spot mediating sucrose binding. Furthermore, this position is structurally equivalent to the SoFase Asn-254, which was reported to be involved in altering the hydrolase/transferase ratio but mostly in modulating the transfructosylating product specificity (24). According to this observation, preliminary analysis shows that the H343T mutant shows an altered profile of FOS production

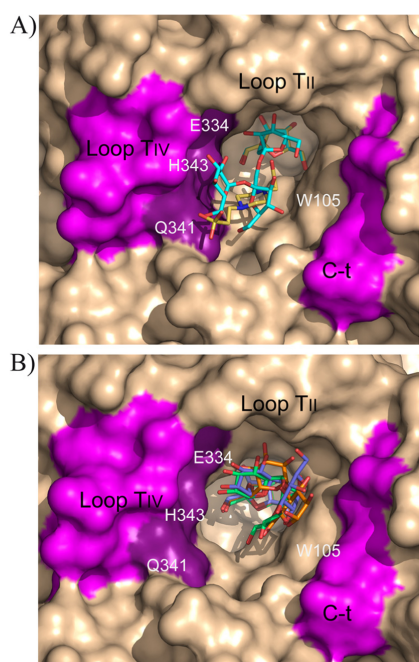


FIGURE 7. Two binding modes of XdINV. Detail of the XdINV active site representing main structural features as follows: subsite +1 is common to all complexes, being shaped by T_{II} and by T_{IV} (light violet, key residues highlighted in dark violet); subsite +2 is created by stacking to Trp-105 and polar interaction to the main chain of the C-terminal region (light violet), but two different sites are observed depending on the substrate. *A*, inulin-type ligands (represented by nystose in cyan and Hepes in yellow) are accommodated by loop T_{IV} . *B*, levan or neo-type ligands are allocated by loop T_{II} (represented by neokestose in green, raffinose in slate, and neoerlose in orange). Additional subsite +3 for inulin-type substrates is formed essentially by loop T_{IV} (*A*).

(data not shown). In contrast, the altered catalytic efficiency observed in the Q341N and N342S mutants might be explained in terms of a different mobility of the Glu-334–His-343 loop, as none of these two residues interact with sucrose. As commented previously, mobility must be a feature of this loop, as it appears highly disordered in the free enzyme being clearly defined upon ligand binding. Consequently, changes at this loop may modify the substrate access to the active site thus altering catalysis. However, the marked impaired activity of the Q341N and N342S mutants on nystose, which is not apparent on raffinose, is consistent with that observed in the complexes and clearly supports the role of these residues and loop T_{IV} in shaping subsites +2 and +3 only for inulin-type substrates.

Finally, a truncated enzyme lacking the last residues comprising the 660 GLFAGY sequence yields a fully inactive enzyme (Table 3). This segment, which is an extended arm following the small helix at the C terminus, surrounds the active site cavity and harbors Gly-660 and Leu-661 shaping subsite +2. Removal of this region probably produces a distortion in the active site architecture of XdINV that is detrimental for activity or even for proper folding of the dimer, as this truncated enzyme is not detected by PAGE (data not shown).

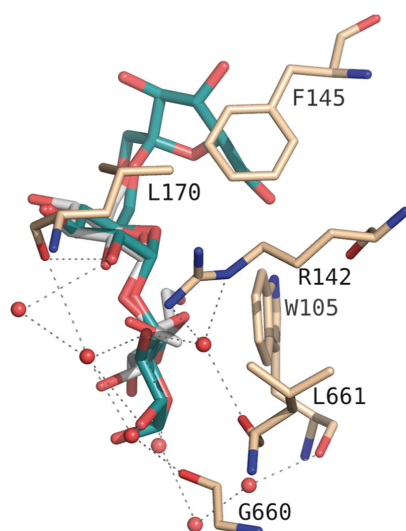
Structural Basis of the Neo/Hetero-FOS Production of XdINV—As commented on previously, 1-kestose and neokestose are the main transglycosylation products obtained with XdINV from sucrose. The experimental complexes here presented show the binding mode of product 1-kestose in the binding site, from which we can infer how the acceptor sucrose would be positioned in subsites +1 and +2 of an enzyme-fructose intermediate complex for the subsequent transfer of this sugar to form a $\beta(2-1)$ -linkage. Thus, the terminal glucose of acceptor sucrose would be mostly oriented through polar interactions to loop T_{IV} , as shown in the schematic representation of the transfructosylation mechanism given in Fig. 8. In contrast, the complex obtained by soaking into neokestose shows this sugar in a donor substrate mode, *i.e.* its Fru- $\beta(2-1)$ -Glu moiety is occupying the $-1/+1$ subsites. Instead, the XdINV-neoerlose complex represents the binding mode of a true neo-type product, with its Fru- $\beta(2-6)$ -Glu moiety spanning the $-1/+1$ subsites, and therefore the binding of acceptor sucrose to produce neokestose can be extrapolated from this complex (Fig. 8A). Thus, although there are also polar interactions stabilizing the hydroxyls, subsite +2 would be mostly determined by stacking of the sugar ring to Trp-105. Interestingly, this trait explains why XdINV is able to fructosylate α -linked glucosides, including maltose, isomaltulose, isomaltose, trehalose, and higher oligosaccharides as maltotriose or maltotetraose, as reported previously (10). In fact, many of these α -glucosides are better acceptors than sucrose, which might be explicated in terms of a more efficient hydrophobic interaction to Trp-105 at subsite +2. In turn, this feature explains why oligosaccharides such as lactose and lactulose did not act as fructose acceptors due to their $\beta(1-4)$ linkage, which is unable to accommodate properly through staking to Trp-105.

The two binding sites described here for acceptor sucrose to form inulin or neo-type products (Fig. 8B) are supported by the observed XdINV-Fru-Hepes and XdINV-Fru-Bistris propane ternary complexes, as shown in Fig. 6. Thus, the Hepes molecule is located similarly to the nystose and mimics a putative donor substrate for a $\beta(2-1)$ transfer to fructose. In contrast, the large and flexible Bistris propane molecule seems apparently able to span both subsites, but in any case, both buffer molecules have a hydroxyl oxygen pointing to the fructose O6 atom, located 3.6 Å away, and thus apparently available to create a linkage. This is also consistent with the altered FOS profile observed in versions of XdINV mutated at loop T_{IV} as commented on previously. Further studies are underway to fully characterize the broad synthetic specificity of this enzyme and to investigate new potential acceptors yielding novel bioconjugates of biomedical interest.

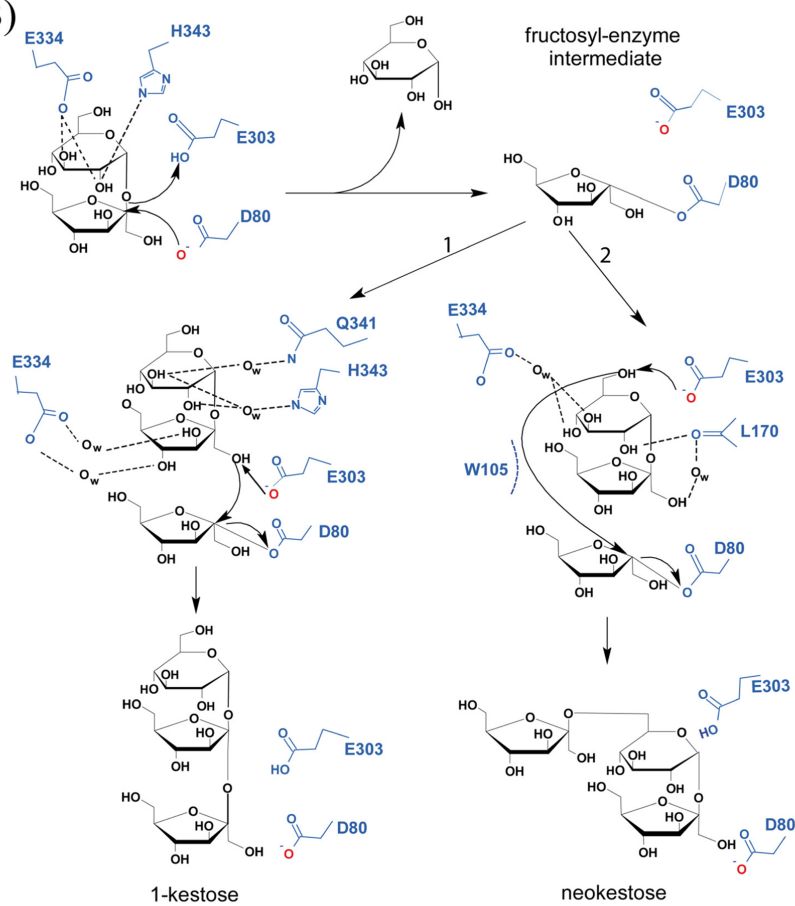
Discussion

Fructans, the fructose-rich polymers biosynthetically derived from sucrose, are important non-structural storage compounds in many bacteria, fungi, and numerous plant species. They are associated with essential biological processes as sucrose metabolism and vegetative development. Also, fructans are biotechnologically relevant compounds due to its prebiotic effect. Consequently, the enzymes involved in fructans and sucrose processing are subject to extensive research with the

A)



B)



aim to uncover their biological role and also to improve their enzymatic efficiency.

In this study, we described the structure of β -fructofuranosidase from *X. dendrorhous* that exhibits novel features regulating activity. The crystal structure shows a bimodular arrangement of its polypeptide chain similar to that described previously for other GH32 enzymes. However, a long 30-residue extension at its C terminus, together with a glycan chain attached to a non-conserved Asn, mediates the formation of an unusual dimer. This dimer shows a peculiar topology that locates both active sites in the same face of the protein and apparently connected by a long crevice. A blast search shows about 60 entries showing 30–60% similarity with XdINV sequence. These sequences include hydrolases or transferases mostly from *Basidiomycota* but also from *Ascomycota* (data not shown). Alignment of the corresponding sequences reveals that they all display both determinants that are shown to be key determinants for the oligomerization state in XdINV, *i.e.* the C-terminal extension and the *N*-glycosylation sequon at a position equivalent to Asn-107, and consequently, the dimer described here would be a model for this group of GH32 enzymes.

The unusual topology of the XdINV dimer suggests the possibility that it might be able to accommodate a putative polymeric substrate as it could be a mixed fructan with branched $\beta(2\text{--}1)/\beta(2\text{--}6)$ linkages (42, 43) that, indeed, is degraded by XdINV. Fructans present in plants are a mixture of heterogeneous oligomers and polymers with a composition depending on the environment and the development state of the plant (38). Key enzymes for plant fructan metabolism include a complex set of biosynthetic and hydrolytic enzymes, including different types of fructan exohydrolases and fructosyltransferases. Consequently, fungi isolated from plant environments may have evolved to interact or degrade these molecules, by expressing their own enzymes. Nevertheless, the possibility of the yeast producing these compounds cannot be discarded, as the biological function of microbial fructans has also been related to the creation of a physical barrier, by enhancing resistance to environmental stresses and assisting in nutrient assimilation (Ref. 43 and references therein). In fact, fructans have proved to act as virulence factors in colonization and to play a critical factor in the pathogenesis of disease-related microorganisms. Many studies have shown that bacteria that synthesize fructans interact with eukaryotic hosts and can be found in symbiotic association with plants. Although the reported studies have been performed in bacteria, it is conceivable that fungi develop a similar mechanism to adapt to the environment. In this sense, it should be noted that the yeast *X. dendrorhous* has been isolated from soil samples collected at different cold ecosystems. The exact physiological role of XdINV and the target substrate at its natural habitat are difficult to envisage, and therefore, it is

hard to derive precise structure-function relationships, but clearly, the peculiar topology of its molecular surface is a remarkable and distinctive trait within GH32 enzymes.

The crystallographic analysis of the XdINV complexes described here gives a detailed picture of the main features of its active site. Previous studies showed that most GH32 enzymes present a shallow pocket with residues shaping a conserved fructose-binding pocket, at subsite -1 , and several hydrophobic and polar interactions defining subsite $+1$. In contrast, the substrate is loosely bound at subsite $+2$ or dangling into the solvent. In agreement with this observation, short substrates such as 1-kestose or raffinose are accommodated in essentially the same conformation in all the formerly described complexes (13, 16, 18, 19). In contrast, XdINV presents a deep cavity showing multiple binding sites, a feature that has been previously described for two other fungal enzymes, the AjFT and the SoFase. The reported complexes of these enzymes with long substrates reveal that the active sites diverge from subsite $+2$, and thus, nystose bound at the AjFT catalytic pocket locates its terminal glucose at a position that is occupied by Trp-105 in XdINV (see Fig. 7A) (16). A very short T_1 loop and the absence of the C-terminal segment make a very different active site pocket in AjFT, which is remarkable taking into account that this enzyme is phylogenetically close to XdINV. Likewise, Frunystose bound at the SoFase accommodates in a precise conformation that mimics the polymeric substrate inulin (24). As commented previously, both yeast enzymes are distinctive within GH32 in being dimers, and both active sites must be considered in terms of their oligomeric state. In the case of SoFase, the topology of the active site seemed designed to optimally fit the conformation of the polymer inulin, which is consistent with its high inulinase activity. Thus, the topology of sites $+2$ and further from the cleaved linkage must reflect the profile of its natural substrate and consequently the physiological role of each enzyme in its environment. Remarkably, XdINV presents both inulin- and levan-type binding sites and an open pocket filled with many water molecules that confer plasticity to the enzyme. Therefore, this plasticity might be consistent with a more complex substrate as are the mixed fructans commented on previously, in contrast to the rigid active site of SoFase adapted to accommodate the more regular inulin.

Carbohydrate-processing enzymes have developed intricate molecular mechanisms to cope with their highly heterogeneous substrates. Understanding these mechanisms will be of general interest to uncover essential biological processes at the molecular level. In the case of XdINV, an insertion at the C-terminal polypeptide and glycosylation have been shown to be essential in oligomerization and thus in shaping a peculiar active site. In turn, the plasticity of its active site makes the enzyme a valuable and flexible biocatalyst to produce novel compounds with bio-

FIGURE 8. Proposed mechanism of XdINV transfructosylating activity. A, view of the XdINV-neoerlose complex, showing the ligand in cyan and the main interactions of its Glu- $\alpha(1\text{--}4)$ Glu moiety located at subsites $+1$ and $+2$. The position of this portion represents a potential binding mode of acceptor maltose for the formation of a $\beta(2\text{--}6)$ linkage of glucose to fructose. The position of a putative sucrose acceptor molecule (white) has been inferred by superposition of its glucose to the glucose found at subsite $+1$ of neoerlose and adjusting manually the fructose unit by small torsion of its glycosidic bond, to best packing to Trp-105 side chain. B, scheme of the proposed mechanism. The donor-substrate sucrose is hydrolyzed by nucleophilic attack forming the fructosyl-XdINV intermediate; a subsequent sucrose molecule (the acceptor substrate) can enter the active site pocket with binding mode 1 generating 1-kestose or, alternatively, with binding mode 2 and generating neokestose.

Structure and Function of Dimeric XdINV

technological relevance. The potential of XdINV to produce novel bioconjugates is underway.

Author Contributions—J. S. A. and M. F. L. conceived and coordinated the study. D. L. performed the production of the native protein. M. G. P., Z. M., and M. F. L. conducted molecular biology assays and kinetics measurements and analyzed the results. M. R. E., B. G., and J. S. A. designed and performed the crystallographic work and interpreted the results. J. S. A. wrote the paper, and all authors read and commented on the manuscript.

Acknowledgments—We thank the staff of the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for providing access and for technical assistance at beamline ID23.1. We also thank the German Electron Synchrotron (Desy) at Hamburg, Germany, for assistance at Petra III beamline, and the Synchrotron Radiation Source at Alba (Barcelona, Spain) for assistance at BL13-XALOC beamline.

References

1. Linde, D., Macias, I., Fernández-Arrojo, L., Plou, F. J., Jiménez, A., and Fernández-Lobato, M. (2009) Molecular and biochemical characterization of a β -fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Appl. Environ. Microbiol.* **75**, 1065–1073
2. Linde, D., Rodríguez-Colinas, B., Estévez, M., Poveda, A., Plou, F. J., and Fernández Lobato, M. (2012) Analysis of neo-fructooligosaccharide production mediated by the extracellular β -fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Bioresour. Technol.* **109**, 123–130
3. Tian, F., and Karboune, S. (2012) Enzymatic synthesis of fructooligosaccharides by levansucrase from *Bacillus amyloliquefaciens*: specificity, kinetics, and product characterization. *J. Mol. Catal. B-Enzym.* **82**, 71–79
4. Rodríguez-Alegria, M. E., Enciso-Rodríguez, A., Ortiz-Soto, M. E., Casani, J., Olvera, C., and Munguia, A. L. (2010) Fructooligosaccharide production by a truncated *Leuconostoc citreum* inulosucrase mutant. *Bio-catal. Biotransfor.* **28**, 51–59
5. Alvaro-Benito, M., de Abreu, M., Fernández-Arrojo, L., Plou, F. J., Jiménez-Barbero, J., Ballesteros, A., Polaina, J., and Fernández-Lobato, M. (2007) Characterization of a β -fructofuranosidase from *Schwanniomyces occidentalis* with transfructosylating activity yielding the prebiotic 6-kestose. *J. Biotechnol.* **132**, 75–81
6. Franco-Robles, E., and Lopez, M. G. (2015) Implication of fructans in health: immunomodulatory and antioxidant mechanisms. *Scientific World J.* **10**, 1155/2015/289267
7. Marx, S. P., Winkler, S., and Hartmeier, W. (2000) Metabolization of β -(2,6)-linked fructose-oligosaccharides by different bifidobacteria. *FEMS Microbiol. Lett.* **182**, 163–169
8. Kilian, S., Kritzinger, S., Rycroft, C., Gibson, G., and du Preez, J. (2002) The effects of the novel bifidogenic trisaccharide, neokestose, on the human colonic microbiota. *World J. Microb. Biotechnol.* **18**, 637–644
9. Lim, J. S., Lee, J. H., Kang, S. W., Park, S. W., and Kim, S. W. (2007) Studies on production and physical properties of neo-FOS produced by co-immobilized *Penicillium citrinum* and neo-fructosyltransferase. *Eur. Food Res. Technol.* **225**, 457–462
10. Gimeno-Pérez, M., Santos-Moriano, P., Fernandez-Arrojo, L., Poveda, A., Jimenez-Barbero, J., Ballesteros, A. O., Fernandez-Lobato, M., and Plou, F. J. (2014) Regioselective synthesis of neo-erlose by the β -fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Process Biochemistry* **49**, 423–429
11. Henrissat, B., Sulzenbacher, G., and Bourne, Y. (2008) Glycosyltransferases, glycoside hydrolases: surprise, surprise! *Curr. Opin. Struct. Biol.* **18**, 527–533
12. Pons, T., Naumoff, D. G., Martínez-Fleites, C., and Hernández, L. (2004) Three acidic residues are at the active site of a β -propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. *Proteins* **54**, 424–432
13. Alberto, F., Jordi, E., Henrissat, B., and Czjzek, M. (2006) Crystal structure of inactivated *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose. *Biochem. J.* **395**, 457–462
14. Bujacz, A., Jedrzejczak-Krzepkowska, M., Bielecki, S., Redzynia, I., and Bujacz, G. (2011) Crystal structures of the apo-form of β -fructofuranosidase from *Bifidobacterium longum* and its complex with fructose. *FEBS J.* **278**, 1728–1744
15. Nagem, R. A., Rojas, A. L., Golubev, A. M., Korneeva, O. S., Eneyskaya, E. V., Kulminkaya, A. A., Neustroev, K. N., and Polikarpov, I. (2004) Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. *J. Mol. Biol.* **344**, 471–480
16. Chuankhayan, P., Hsieh, C. Y., Huang, Y. C., Hsieh, Y. Y., Guan, H. H., Hsieh, Y. C., Tien, Y. C., Chen, C. D., Chiang, C. M., and Chen, C. J. (2010) Crystal structures of *Aspergillus japonicus* fructosyltransferase complex with donor/acceptor substrates reveal complete subsites in the active site for catalysis. *J. Biol. Chem.* **285**, 23251–23264
17. Pouyez, J., Mayard, A., Vandamme, A. M., Roussel, G., Perpète, E. A., Wouters, J., Housen, I., and Michaux, C. (2012) First crystal structure of an endo-inulinase, INU2, from *Aspergillus ficuum*: discovery of an extra-pocket in the catalytic domain responsible for its endo-activity. *Biochimie* **94**, 2423–2430
18. Verhaest, M., Lammens, W., Le Roy, K., De Ranter, C. J., Van Laere, A., Rabijns, A., and Van den Ende, W. (2007) Insights into the fine architecture of the active site of chicory fructan 1-exohydrolase: 1-kestose as substrate vs sucrose as inhibitor. *New Phytol.* **174**, 90–100
19. Lammens, W., Le Roy, K., Yuan, S., Vergauwen, R., Rabijns, A., Van Laere, A., Strelkov, S. V., and Van den Ende, W. (2012) Crystal structure of 6-SST/6-SFT from *Pachysandra terminalis*, a plant fructan biosynthesizing enzyme in complex with its acceptor substrate 6-kestose. *Plant J.* **70**, 205–219
20. Verhaest, M., Lammens, W., Le Roy, K., De Coninck, B., De Ranter, C. J., Van Laere, A., Van den Ende, W., and Rabijns, A. (2006) X-ray diffraction structure of a cell-wall invertase from *Arabidopsis thaliana*. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1555–1563
21. Alvaro-Benito, M., Polo, A., González, B., Fernández-Lobato, M., and Sanz-Aparicio, J. (2010) Structural and kinetic analysis of *Schwanniomyces occidentalis* invertase reveals a new oligomerization pattern and the role of its supplementary domain in substrate binding. *J. Biol. Chem.* **285**, 13930–13941
22. Sainz-Polo, M. A., Ramírez-Escudero, M., Lafraya, A., González, B., Marín-Navarro, J., Polaina, J., and Sanz-Aparicio, J. (2013) Three-dimensional structure of *Saccharomyces invertase*: role of a non-catalytic domain in oligomerization and substrate specificity. *J. Biol. Chem.* **288**, 9755–9766
23. Trollope, K. M., van Wyk, N., Kotjomela, M. A., and Volschenk, H. (2015) Sequence and structure-based prediction of fructosyltransferase activity for functional subclassification of fungal GH32 enzymes. *FEBS J.* **282**, 4782–4796
24. Alvaro-Benito, M., Sainz-Polo, M. A., González-Pérez, D., González, B., Plou, F. J., Fernández-Lobato, M., and Sanz-Aparicio, J. (2012) Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β -fructofuranosidase, an enzyme that produces prebiotics. *J. Biol. Chem.* **287**, 19674–19686
25. Gimeno-Pérez, M., Linde, D., Fernández-Arrojo, L., Plou, F. J., and Fernández-Lobato, M. (2015) Heterologous overproduction of β -fructofuranosidase from yeast *Xanthophyllomyces dendrorhous*, an enzyme producing prebiotic sugars. *Appl. Microbiol. Biotechnol.* **99**, 3459–3467
26. Polo, A., Linde, D., Estévez, M., Fernández-Lobato, M., and Sanz-Aparicio, J. (2010) Crystallization and preliminary x-ray diffraction analysis of the fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1441–1444
27. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281
28. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst.*

- tallogr. D Biol. Crystallogr.* **67**, 235–242
29. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
 30. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025
 31. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643
 32. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
 33. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
 34. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655
 35. Schüttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363
 36. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291
 37. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, version 1.6.0.0, DeLano Scientific, San Carlos, CA
 38. Ritsema, T., and Smeekens, S. (2003) Fructans: beneficial for plants and humans. *Curr. Opin. Plant Biol.* **6**, 223–230
 39. Lopez, M. G., Mancilla-Margalli, N. A., and Mendoza-Diaz, G. (2003) Molecular structures of fructans from *Agave tequilana* Weber var. azul. *J. Agric. Food Chem.* **51**, 7835–7840
 40. Lammens, W., Le Roy, K., Van Laere, A., Rabijns, A., and Van den Ende, W. (2008) Crystal structures of *Arabidopsis thaliana* cell-wall invertase mutants in complex with sucrose. *J. Mol. Biol.* **377**, 378–385
 41. Le Roy, K., Lammens, W., Verhaest, M., De Coninck, B., Rabijns, A., Van Laere, A., and Van den Ende, W. (2007) Unraveling the difference between invertases and fructan exohydrolases: a single amino acid (Asp-239) substitution transforms *Arabidopsis cell* wall invertase1 into a fructan 1-exohydrolase. *Plant Physiol.* **145**, 616–625
 42. Cimini, S., Locato, V., Vergauwen, R., Paradiso, A., Cecchini, C., Vandenoel, L., Verspreet, J., Courtin, C. M., D'Egidio, M. G., Van den Ende, W., and De Gara, L. (2015) Fructan biosynthesis and degradation as part of plant metabolism controlling sugar fluxes during durum wheat kernel maturation. *Front. Plant Sci.* **6**, 89
 43. Velázquez-Hernández, M. L., Baizabal-Aguirre, V. M., Bravo-Patiño, A., Cajero-Juárez, M., Chávez-Moctezuma, M. P., and Valdez-Alarcón, J. J. (2009) Microbial fructosyltransferases and the role of fructans. *J. Appl. Microbiol.* **106**, 1763–1778
 44. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
 45. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–324

3.4) MATERIAL INÉDITO

Con el objetivo de avanzar en el conocimiento del mecanismo de acción de ScInv, se intentaron obtener cristales de mejor calidad para aumentar la resolución de los datos de difracción. Sin embargo, tras muchos intentos y la aplicación de distintas técnicas de cristalización, los mejores datos de difracción son los reportados en el primer trabajo de investigación (3,3 Å). Desafortunadamente, esta resolución no es adecuada para el análisis detallado de complejos enzima-sustrato.

En el caso de la XdInv sin embargo, sí se analizaron diversos complejos enzima-ligando. Estos resultados mostraron que el centro activo de XdInv tiene dos sitios de unión a sustrato, uno polar dirigido hacia el bucle T_{IV}, y otro hidrofóbico, definido esencialmente por el Trp105 y el extremo C-terminal. Además, se encontraron dos complejos ternarios (Fru-HEPES y Fru-BTpro), donde las moléculas de HEPES y Bis-Tris propano mimetizan la posición de un posible sustrato aceptor. Estos resultados resaltan la amplia especificidad del centro activo de XdInv y, por tanto, su potencial versatilidad en la síntesis de nuevos fructo-conjugados. Además, los primeros análisis de los cristales obtenidos de la proteína nativa mostraron la presencia de un ligando cuya existencia sólo se podía explicar como un hipotético producto de fructosilación de la maltosa procedente del medio de cultivo. Todo ello sugirió una serie de experimentos dirigidos a explorar posibles aceptores que condujeran a la obtención de neo-FOS y otros novedosos compuestos bioactivos.

Así, en primer lugar, se intentó cristalizar el complejo ternario entre el donador fructosa y distintos compuestos que demostraron actuar como aceptores³⁵ como la isomaltulosa, la maltotetraosa, la maltotriosa, la maltosa, la isomaltosa, la rafinosa y la trehalosa. En ninguno de los casos se obtuvo el complejo ternario (fructosa + sustrato aceptor). Para evitar la interferencia de moléculas como HEPES y Bis-Tris propano, se optimizaron las condiciones de cristalización para obtener cristales en ausencia de tampón. Así, se crecieron los cristales en menor cantidad de citrato de sodio a pH neutro, y en ausencia de solución tamponadora (1,1-1,2 M citrato de sodio pH 6,8), y se repitieron los ensayos de *soaking*. Primero, los cristales fueron pre-incubados en solución precipitante suplementada con 20 mM fructosa. Luego, los cristales se sumergieron en solución precipitante suplementada con 50-150 mM de los distintos aceptores antes enunciados. Asimismo, se intentaron obtener los complejos por co-cristalización de la proteína con fructosa y los distintos aceptores. Sin embargo, ninguno de estos experimentos condujo a la obtención de complejos ternarios.

Por otra parte, estudiamos el papel de otras moléculas que no funcionaban como aceptores pero que, en algunos casos, mostraron actividad inhibitoria³⁵. Para ello, se realizaron ensayos de *soaking* con 50-100 mM celobiosa o lactulosa, obteniendo sólo el complejo XdInvD80A-lactulosa (figura 3.1). La lactulosa se sitúa desplazada de la forma canónica de los sustratos (definido por la sacarosa). Concretamente, el residuo de fructosa de la lactulosa se sitúa en la posición que ocupa el enlace glicosídico de la sacarosa, y la unidad de glucosa se sitúa mediante interacción hidrofóbica por apilamiento con Trp105. Además, la lactulosa se estabiliza con nuevos contactos con residuos que no estaban implicados en el reconocimiento de sustratos ni productos (Gln97, Asn79 y Arg220), y no interacciona con His343, ni con el bucle T_{II} o el extremo C-terminal. Su unión bloquea la entrada de sustratos en el centro activo, explicando la actividad de la lactulosa como inhibidor de la actividad de XdInv.

Asimismo, se intentó estudiar si XdInv presenta subsitios de unión adicionales a los caracterizados en los primeros complejos descritos. Para ello se incubaron los cristales con polímeros como inulina, levano o fructano de ágave. Sin embargo, en el centro activo sólo se encontró una molécula de sacarosa, fructosa o 1-kestosa, respectivamente. Los polímeros comerciales contienen

pequeñas trazas de subproductos más cortos que, claramente, son más afines que los polisacáridos en el centro activo de XdInv. Alternativamente, se obtuvo el complejo XdInvD80A con fructosil-nistosa (GF4).

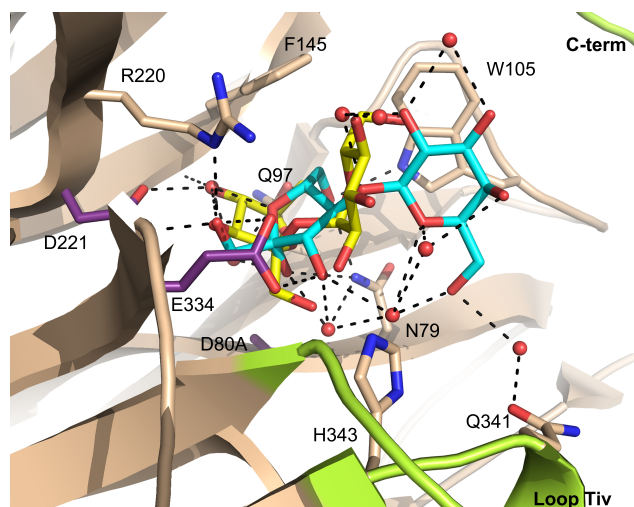


Figura 3.1. Estructura de XdInv en complejo con sacarosa (amarillo) y lactulosa (azul). Vista del centro activo donde se muestran los residuos más relevantes con los que el inhibidor lactulosa interacciona. En verde se resaltan los bucles implicados en la especificidad de sustrato de XdInv, y en morado los residuos catalíticos.

En la figura 3.2 se muestra la disposición de este ligando en el centro activo de XdInv, en comparación con la disposición de la nistosa. Como se puede observar, la conformación de ambos sustratos así como la disposición de sus subunidades en los subsitios +3 y +4 es distinta. Sin embargo, es interesante que los dos sustratos se acomodan esencialmente por interacción con el bucle T_{IV} , lo cual confirma que este sitio polar es el sitio de unión de sustratos tipo inulina o ^1F-FOS . Además, el anillo terminal del extremo reductor de ambos sustratos se sitúa en una posición muy similar. Esto parece indicar que XdInv podría tener cuatro subsitios de unión frente al polímero inulina, aunque la plasticidad de su bolsillo catalítico permite la existencia de más subsitios para acomodar oligosacáridos de tamaño medio, como la fructosil-nistosa, que presentan libertad conformational.

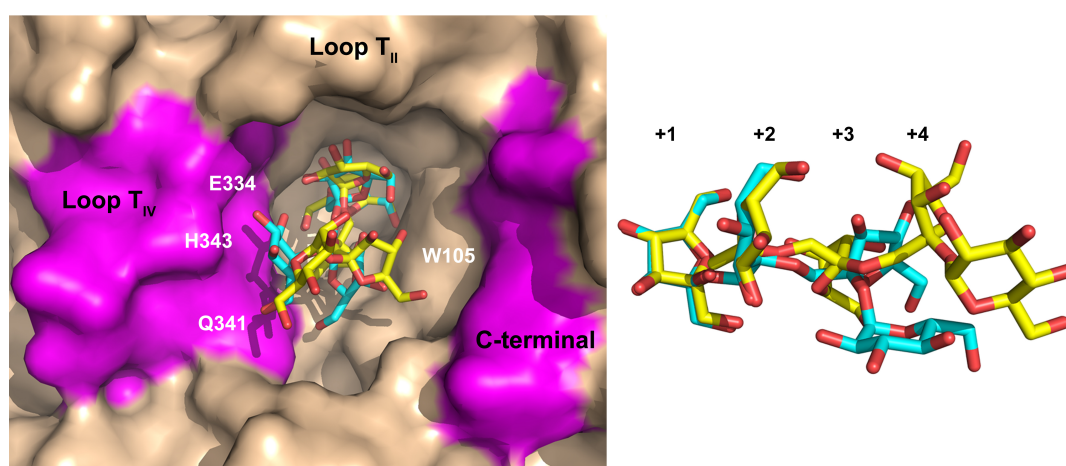


Figura 3.2. Comparación de la disposición de la nistosa y la fructosil-nistosa en el centro activo de XdInv. A la izquierda, se muestra una vista frontal de la cavidad de XdInv con ambos sustratos en el centro catalítico. En rosa se indican el bucle T_{IV} (izda.) y el C-terminal de la enzima (dcha.). A la derecha, se presenta una vista lateral de la conformación de la fructosil-nistosa (amarillo) y la nistosa (azul).

Por último, se están llevando a cabo otros experimentos de cristalización para explorar nuevas capacidades de transfructosilación de XdInv. Para ello, se han realizado *soakings* de los cristales en la solución precipitante suplementada con fructosa y distintos compuestos como polialcoholes (xilitol y sorbitol), antibióticos (estreptomicina, kanamicina, neomicina y tobramicina) o antioxidantes (hidroquinona, resveratrol, catecol, EGCG, procianidina, quercetina, hidroxitirosol y *p*-nitrofenol). Hasta la fecha, sólo se han obtenido complejos ternarios con algunos fenoles antioxidantes, como la hidroquinona y el catecol. Datos preliminares muestran que, aunque ambas moléculas se posicionan mediante interacción por apilamiento con Trp105, el catecol actúa como un inhibidor de XdInv, mientras que la hidroquinona es sustrato aceptor. Por otra parte, dada su similitud molecular con el HEPES, se seleccionó el hidroxitirosol como un posible sustrato aceptor. Es interesante destacar que los ensayos preliminares confirman esta hipótesis y actualmente hemos obtenido el complejo de XdInv con este glicoderivado. Sorprendentemente, el fragmento del hidroxitirosol se estabiliza únicamente por interacciones con el bucle T_{II}. Actualmente, estamos realizando ensayos de cristalización con otros antioxidantes, además de pruebas de caracterización bioquímica para completar estos estudios (figura 3.3).

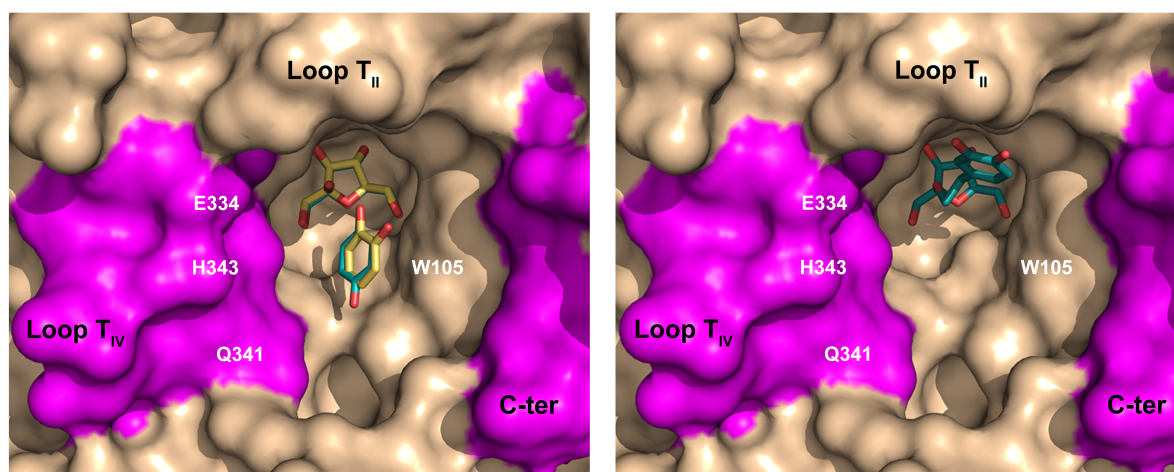


Figura 3.3. Comparación de la disposición de los distintos antioxidantes en el centro activo de XdInv. A la izquierda, los complejos ternarios fructosa y catecol (en amarillo) y fructosa e hidroquinona (en azul). A la derecha, el producto fructosil-hidroxitirosol. En rosa se indican el bucle T_{IV} y el segmento C-terminal de la enzima.

Simultáneamente, se están realizando experimentos con mutantes puntuales del bucle T_{IV} de la enzima XdInv, diseñados con el objetivo de intentar incrementar la actividad transfructosilasa para la producción de FOS, su especificidad por determinados productos, así como mejorar la síntesis de bioconjugados (figuras 3.1 y 3.3).

En paralelo a los ensayos de producción de bioconjugados con XdInv también se ha trabajado con la β -fructofuranosidasa de *Schwanniomyces occidentalis* (SoFfase) para la síntesis de fructoconjugados, principalmente, la síntesis de fructosil-alditales. Al igual que procedimos con la XdInv, se intentó obtener los complejos ternarios con fructosa y distintos alditales, como manitol, sobitol y xilitol, así como con los disacáridos trehalosa y lactosa, pero no se obtuvieron resultados. Recientemente se ha obtenido el complejo SoFfase con el producto de transfructosilación fructosil-eritritol (figura 3.4). Próximos experimentos de cristalización proporcionarán nuevos datos sobre la capacidad de sintetizar bioconjugados por parte de esta enzima.

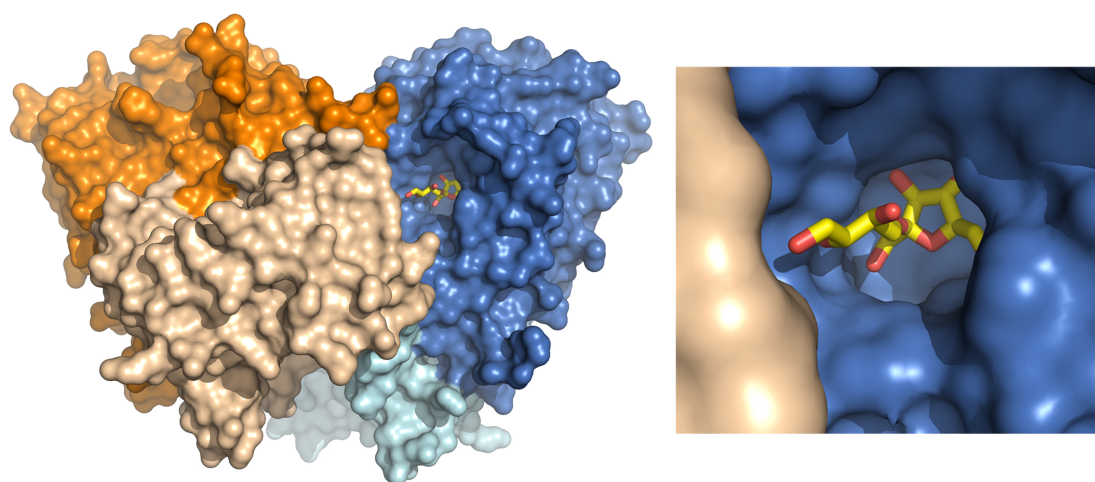


Figura 3.4. La β -fructofuranosidasa de *Schwanniomyces occidentalis* en complejo con fructosil-eritritol. En azul se muestra el monómero A y en naranja el monómero B (color oscuro, dominios catalítico; color claro, β -sándwich). A la derecha se muestra una vista ampliada del centro activo. En amarillo se representa una molécula de fructosil-eritritol.

3.5) DISCUSIÓN

Los trabajos aquí presentados proporcionan información muy útil para entender mejor las bases moleculares de la actividad de las enzimas que modifican fructanos. En general, puede considerarse que la especificidad de sustrato de los miembros de la familia GH32 está regulada a tres niveles¹⁰².

El primer nivel está determinado por la secuencia de aminoácidos. Por una parte, aunque los monómeros son similares en sus estructuras secundarias, muestran diferencias en la longitud y conformación de los bucles que constituyen el centro activo, lo que conlleva cambios en la anchura y profundidad de los bolsillos catalíticos⁶². Es por ello que, dentro de la familia GH32 existen endo- y exo-hidrolasas (figura 3.5.A). Por otra parte, los alineamientos de secuencia de los distintos miembros GH32 revelan la baja identidad de estos bucles del centro activo, exhibiendo distinta distribución de carga en la cavidad catalítica¹⁰². Estos cambios determinan el tipo de actividad de la enzima. En la figura 3.5.B se muestran las diferencias en carga, anchura y profundidad de los bolsillos catalíticos de las enzimas GH32 de levadura.

Por otra parte, mutaciones puntuales en los subsitios +1 y/o +2 de determinados residuos, pueden promover productos únicos o, incluso, nuevas especificidades. Un ejemplo claro son las mutaciones llevada a cabo en SoFfase por nuestro grupo, a partir del análisis de complejos de la enzima con sus sustratos y/o productos¹⁰¹ (figura 3.6). Mientras el mutante Q228V produce específicamente 6-kestosa, la mutación N254T genera una mezcla de productos entre los que se encuentra la neokestosa, un producto no descrito con anterioridad para esta enzima. Por tanto, el análisis estructural permite identificar residuos esenciales cuyo papel puede pasar desapercibido en el análisis meramente secuencial. En este trabajo hemos identificado una serie de residuos claves en el sitio activo de las dos enzimas estudiadas, ScInv y XdInv, y se han diseñado nuevos experimentos para mejorar sus capacidades biosintéticas.

El segundo nivel de control, especialmente para sustratos largos, es la glicosilación. Verhaes M. y colaboradores⁵⁹ reportaron que la incapacidad para degradar inulina de ciertas enzimas de plantas posiblemente se debe a la existencia de una cadena de N-glicanos que bloquea el acceso al centro activo¹⁰³. Hasta el momento, este nivel de regulación sólo se ha descrito para la fructán-exohidrolasa de *Cichorium intybus*.

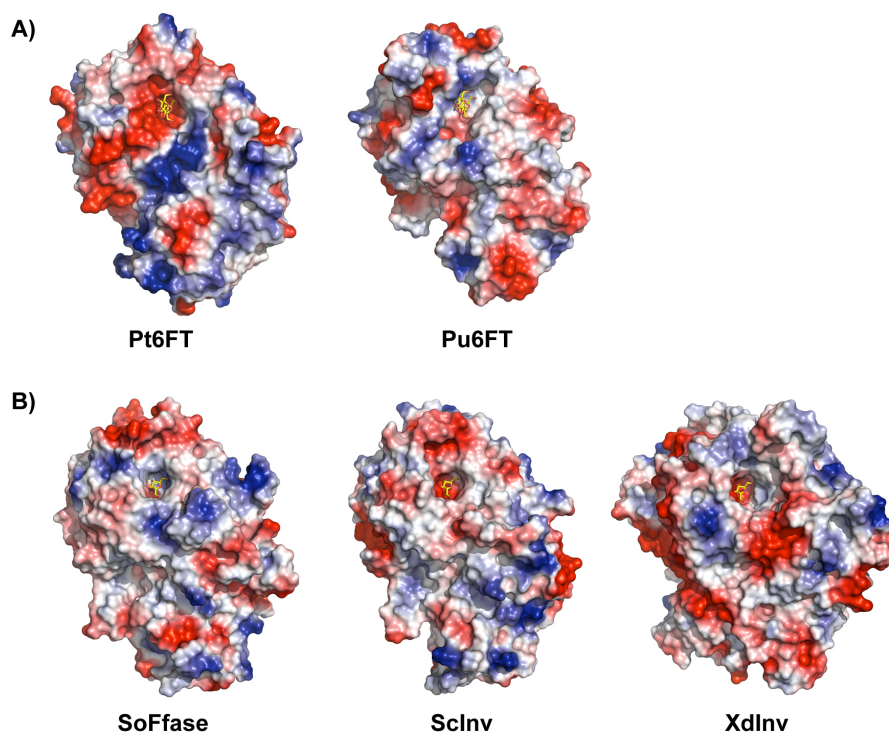


Figura 3.3. Comparación de las estructuras resueltas de la familia GH32. Se muestra la superficie molecular de cada enzima, coloreada según el potencial electrostático (negativo en rojo y positivo en azul). (A) Comparación entre una exo- (izda.) y una endo- (dcha.) fructosil-transferasa. En amarillo se representa una molécula de sacarosa. (B) Comparación de las estructuras de levadura GH32. En el centro activo se muestra una unidad de fructosa en el subsitio -1. Pt6FT (6-exo-fructosiltransferasa de *Pachysandra terminalis*), Pu6FT (6-endo-fructosiltransferasa de *Paenarthrobacter urefaciens*), SoFfase (β -fructofuranosidasa de *Schwanniomyces occidentalis*), Sclnv (invertasa de *Saccharomyces cerevisiae*) y Xdlnv (β -fructofuranosidasa de *Xanthophyllomyces dendrorhous*).

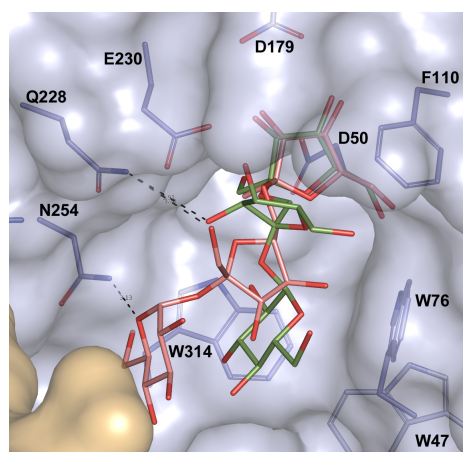


Figura 3.6. Centro activo de SoFfase en complejo con 1-kestosa (verde) y 6-kestosa (rosa). En azul se representa el dominio β -propeller y en beige el dominio β -sándwich. Imagen tomada de ¹⁰¹.

Hay que destacar que las N-glicosilaciones tienen un papel muy importante en la estabilidad estructural de las enzimas¹⁰⁴. En el caso de Xdlnv, la N-glicosilación de la Asn107 es vital para la actividad de la enzima ya que su eliminación produce una enzima inactiva. Sin embargo, los azúcares que componen las N-glicosilaciones de Xdlnv no intervienen directamente en la regulación de la especificidad de sustrato y la inactivación se debe a la destrucción del dímero y, por tanto, de su estructura nativa.

Por último, hemos encontrado en nuestras proteínas de levadura un tercer nivel de regulación de la especificidad basado en el estado oligomérico de la enzima. Hasta hace muy poco, todas las estructuras GH32 reportadas eran monoméricas, pero en 2010 nuestro grupo publicó la estructura de SoFfase, la primera estructura dimérica de la familia GH32⁶⁰. Este estudio mostró, por primera vez, el papel que juega el β -sándwich en la oligomerización, y ésta, a su vez, en la actividad de la enzima. Así, el centro activo está formado por ambos monómeros: el bolsillo catalítico de una subunidad está rodeado por los bucles del dominio β -sándwich de la otra. Además, mutaciones de ciertos residuos de estos bucles produjeron una disminución de la eficiencia catalítica frente a inulina, lo que destaca el papel esencial que tienen la oligomerización en la especificidad, y asigna directamente un papel catalítico al dominio β -sándwich del cuál, como se señaló anteriormente, no se conocía su función dentro de la familia GH32.

Los dos trabajos incluidos en esta Tesis, muestran también oligómeros con un papel directo en la actividad. Parece, por tanto, que este tercer nivel de control de la especificidad de sustrato puede ser característico de las GH32 de levaduras (figura 3.7).

Este tercer nivel de control es tan sofisticado que permite que dos dímeros estructuralmente muy similares, como los que presentan SoFfase y ScInv, sin embargo desarrollen actividades distintas. Así, mientras ScInv sólo presenta actividad invertasa (es activa fundamentalmente frente a sacarosa, o trisacáridos en menor grado), SoFfase además presenta una elevada actividad inulinasa y degrada el polímero³⁴. SoFfase presenta cinco subsitios de unión que parecen adaptados específicamente para albergar la molécula helicoidal de inulina (figura 3.7), mientras que ScInv presenta sólo 2-3 subsitios por oclusión del centro activo en el octámero. Por su parte, XdInv tiene ambos centros activos situados en la misma cara del dímero, mostrando una hendidura que conecta ambos bolsillos catalíticos y que podría sugerir que ambos actúen de forma concertada reconociendo polímeros mixtos ramificados.

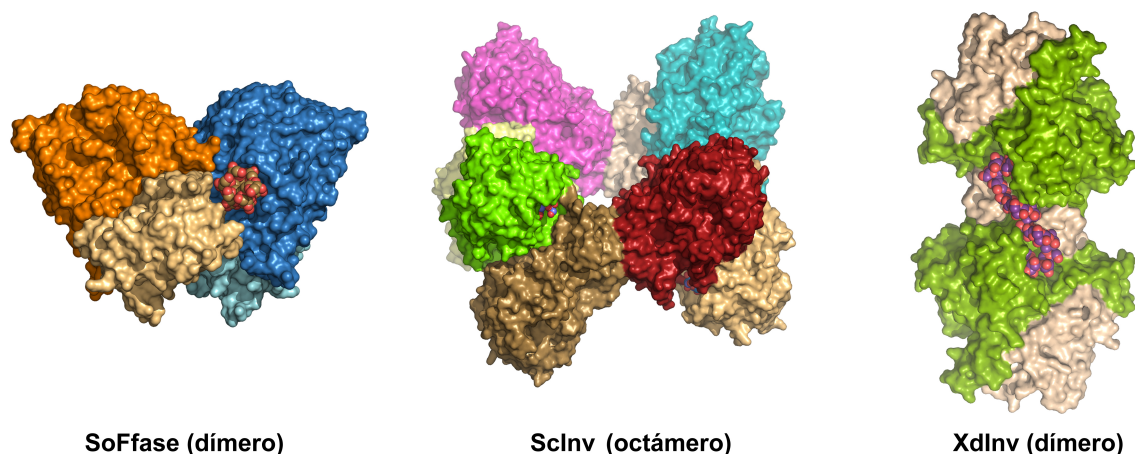


Figura 3.7. Estructura de las enzimas GH32 oligoméricas resueltas hasta la fecha. En los centros activos se observan las especificidades de sustrato proporcionadas por la oligomerización: SoFfase en complejo con inulina resalta su actividad inulinasa, ScInv en complejo con 1-kestosa muestra la especificidad por sustratos cortos de la invertasa de *S. cerevisiae*, y XdInv en complejo con graminano destaca la posibilidad de este nuevo tipo de dímero de alojar sustratos mixtos ramificados.

En resumen, estos dos trabajos de investigación aportan valiosa información acerca del mecanismo de acción de estas enzimas a nivel molecular. Así mismo revelan que, además de la secuencia de aminoácidos de cada proteína, el dominio β -sándwich juega un papel fundamental en la oligomerización de las enzimas GH32 de levadura, y esta oligomerización en la modulación de la especificidad de sustrato. Además del interés básico para entender mejor este grupo de enzimas, esta información es esencial para modular su actividad y potenciar su uso biosintético.

La producción enzimática de FOS es un tema de gran relevancia para la elaboración de nuevos productos que actúen como drogas o ingredientes funcionales. Es por ello que se están buscando nuevas enzimas con actividad optimizada para la producción específica de FOS, y para la síntesis, a nivel industrial, de estos compuestos. Este trabajo incluye dos enzimas capaces de producir β (2-6)-FOS (6 F-FOS y neo-FOS) con mejores propiedades prebióticas que los 1 F-FOS comercializados³⁶.

Por un lado, aunque ScInv tiene una marcada actividad invertasa, es capaz de producir 6-kestosa. Recientemente, Polaina J. y colaboradores¹⁰⁵ diseñaron un método de producción de 6-kestosa a partir de ScInvN21S, una versión mejorada por ingeniería genética de la invertasa de *S. cerevisiae* aquí descrita, que produce el mayor rendimiento de 6-kestosa reportado hasta el momento¹⁰⁶, con la ventaja añadida de eludir la purificación de la enzima. La mutación N21S afecta al motivo del nucleófilo (WMNDPN). Mutaciones en esta secuencia conllevan cambios en la especificidad de sustrato y/o en el ratio hidrólisis:transfructosilación.

Por su parte, XdInv tiene el mayor rendimiento de producción de neokestosa conseguido hasta la fecha¹⁰⁷. Además, el análisis estructural de la enzima mostró la extraordinaria plasticidad de su centro activo y la presencia de un inesperado producto de transfructosilación (neoerlosa) a partir de maltosa como sustrato aceptor, lo que sugirió la posibilidad del uso de XdInv para producir hetero-FOS³⁵. Aunque otras enzimas también han sido descritas con capacidad de producir estos hetero-FOS^{97, 108}, XdInv es capaz de fructosilar distintos glucósidos, como la maltosa, la isomaltulosa y la isomaltosa, además de oligómeros más largos como la maltotriosa y la maltotetraosa³⁵. Concretamente, XdInv es capaz de producir neoerlosa y, pese a que éste no es el mayor rendimiento descrito hasta la fecha⁹⁹, la regioselectividad de la XdInv en la producción de neoerlosa (88% del total de FOS)³⁵ permite una purificación mucho más simple y, por tanto, un proceso más rentable y eficiente.

Tal y como se ha comentado en el apartado de material inédito de este capítulo, se está explorando la capacidad de las GH32 de levadura de fructosilar otros compuestos, como alditoles, antibióticos o antioxidantes, para producir bio-conjugados. Desde hace años se conocen los beneficios farmacocinéticos, y farmacodinámicos, incluyendo la mejora de la solubilidad y de la biodisponibilidad, de la glicosilación de distintos compuestos químicos¹⁰⁹. De forma análoga, la fructosilación de estos compuestos aumentaría su solubilidad, así como sus propiedades funcionales.

En comparación, el centro activo de XdInv es más amplio que el de SoFfase (figura 3.3.B), y la presencia de dos zonas de unión de sustratos (tipo inulina vs. tipo levano) podría permitir la unión de aceptores de diversa topología. De hecho, mientras que la hidroquinona interacciona hidrofóticamente con Trp105, la parte fenólica del biocompuesto fructosil-hidroxitirosol, establece contactos directos con el bucle T_{II}. Por tanto, parece que el centro activo de XdInv es tan versátil que permite la interacción con compuestos de formas muy diversas, acomodando moléculas con estructuras distintas. Sin embargo, la cavidad catalítica más constreñida de SoFfase parece facilitar la estabilización de alditoles frente a otros compuestos más voluminosos, como los antibióticos.

En conclusión, el estudio estructural y la comparación de los centros activos de las enzimas descritas en este trabajo muestran los determinantes esenciales de su distinta especificidad. Además, el análisis estructural ha revelado la amplia especificidad y plasticidad de XdInv, la cuál es capaz de acomodar una amplia variedad de sustratos aceptores, como maltooligosacáridos, polifenoles, etc. La combinación de los datos estructurales y bioquímicos, y los estudios de mutagénesis, proporcionan las claves que permitirán modular la especificidad con el fin de mejorar sus capacidades biosintéticas.

4. ENZIMAS PRODUCTORAS DE IMOS:

α -Glucosidasa de *X. dendrorhous*

4.1) INTRODUCCIÓN

Los isomaltooligosacáridos (IMOS) comercializados en la actualidad son producidos por α -transglucosidasas (EC 2.4.1.24) fúngicas⁴¹, que generan siropes de IMOS, productos heterogéneos en composición, grado de polimerización y eficiencia prebiótica. Por ello, las α -glucosidasas (EC 3.2.1.20) se presentan como una alternativa más eficiente, ya que generan productos más homogéneos.

Las α -glucosidasas se incluyen en cinco familias: GH4, GH13, GH31, GH63 y GH97, aunque mayoritariamente pertenecen a las familias GH13 y GH31⁷⁴, que están incluidas en los clanes GH-D y GH-H, respectivamente. Además de la familia GH a la que pertenecen, las α -glucosidasas se pueden catalogar en función de la especificidad de sustrato, dividiéndose en tres tipos. Las tipo I, reconocen varios tipos de enlaces distintos (heterósidos, como la sacarosa), mientras los tipos II y III prefieren carbohidratos con el mismo tipo (holósidos, como los maltooligosacáridos o los α -glucanos). Así, mientras las tipo I pertenecen a la familia GH13, las α -glucosidasas tipo II y III son miembros de la familia GH31⁷⁴. La diferencia entre el tipo II y III reside en su distinta capacidad de hidrolizar polisacáridos: baja en el tipo II y alta en las del tipo III¹¹⁰.

En comparación a la familia GH13, la familia GH31 ha recibido relativamente poca atención, a pesar de contener enzimas de los tres dominios de la vida (arqueas, bacterias y eucariotas) y de su importancia en diferentes procesos biológicos. Entre ellas, la α -glucosidasa lisosomal humana, cuya deficiencia resulta en la enfermedad de Pompe (que provoca una acumulación de glucógeno en las células musculares); la α -glucosidasa II del retículo endoplasmático, que juega un papel fundamental en el procesado y plegamiento de las glicoproteínas; o las sucrosa-isomaltasa y la maltosa-glucoamilasa humanas, responsables de la degradación final del almidón en el intestino y dianas de inhibición de los medicamentos anti-diabetes como la acarbosa y el miglitol⁶⁷. Además, la familia GH31 también contiene α -xilosidasas, isomaltosil-transferasas y α -glucan-lisas. Recientemente se han caracterizado nuevas actividades dentro de la familia, como la α -galactosidasa de *Pseudobacter saltans*⁷⁰ o la sulfoquinovosidasa de *E. coli*⁷¹.

Hasta la fecha de escritura de este manuscrito, se han resuelto 19 estructuras tridimensionales de la familia GH31 (www.cazy.org/GH31_structure.html), de las que siete son α -glucosidasas, pero ninguna de levaduras. Estos estudios han mostrado la gran complejidad estructural que presentan estas enzimas.

Molecular Characterization and Heterologous Expression of a *Xanthophyllomyces dendrorhous* α -Glucosidase with Potential for Prebiotics Production.

Patricia Gutiérrez-Alonso, María Gimeno-Pérez, Mercedes Ramírez-Escudero, Francisco J. Plou, Julia Sanz-Aparicio and María Fernández-Lobato.

Applied Microbiology and Biotechnology, Vol. 100, Iss. 7, pp. 3125–93135,

1 de Abril del 2016

DOI: 10.1007/s00253-015-7171-3

Interés de la investigación:

La α -glucosidasa de *Xanthophyllomyces dendrorhous* (Xd- α Glu) es una exo-glucosidasa con capacidad de hidrolizar exclusivamente enlaces $\alpha(1-4)$, si bien puede degradar malto-oligosacáridos de distinto grado de polimerización (maltosa, maltotriosa y también maltoheptaosa), e incluso almidón soluble.

Xd- α Glu es una glicoproteína de 990 residuos cuya porción de carbohidratos supone el 12% de su peso molecular (7 kDa aproximadamente). Así, Xd- α Glu es una de las mayores α -glucosidasas monomérica de levadura descrita. Para ahondar en las características propias de esta enzima, es necesario analizar su estructura en el contexto de las otras enzimas caracterizadas y/o resueltas hasta la fecha.

Hay que destacar que Xd- α Glu es capaz de producir isomaltooligosacáridos con enlaces $\alpha(1-6)$, de mayor interés industrial ya que presentan mejores propiedades funcionales y prebióticas. Además, mientras que la mayoría de α -glucosidasas generan disacáridos como productos de reacción, Xd- α Glu genera tri- y tetrasacáridos, con mayor resistencia a la degradación intestinal y, por tanto, mayor capacidad prebiótica. Además, ninguna otra enzima produce los tetrasacáridos aquí descritos (6-O- α -glucosil-maltotriosa y 6-O- α -isomaltosil-maltosa). Por ello, su caracterización estructural puede proporcionar las claves de los determinantes moleculares involucrados en esta especificidad de producto.

Los objetivos de este trabajo fueron:

- Aislar y caracterizar el gen de Xd- α Glu.
- Analizar su secuencia de aminoácidos en el contexto del resto de α -glucosidasas descritas.
- Estudiar los componentes estructurales que determinan la especificidad de sustrato y producto de esta enzima.
- Identificar los residuos catalíticos de Xd- α Glu, así como otros motivos de secuencia característicos de la especificidad de la actividad de GH31 subgrupo 1.
- Caracterizar su expresión heteróloga en el sistema *S. cerevisiae*.

Las conclusiones de esta investigación son:

- La secuenciación de la enzima Xd- α Glu da lugar a ORF (marco de lectura abierto, *Open Reading Frame*) de 2970 bp (pares de bases) que contiene 13 intrones y 14 exones.
- Este gen codifica para una proteína de 990 residuos (110 kDa aprox.) cuya secuencia presenta una elevada similitud con las α -glucosidasas de ascomicetos, como *S. occidentalis*, *S. pombe* y *Aspergillus* sp. Por tanto, es un miembro de la familia GH31.
- El modelo estructural por homología muestra que Xd- α Glu tiene la mayoría de los elementos que caracterizan a los miembros de la familia GH31: los cuatro dominios básicos y los dos bucles característicos del dominio catalítico, típicos de las α -glucosidasas de eucariotas. Sin embargo, Xd- α Glu presenta dos características estructurales novedosas: un inserto en el dominio N-terminal (denominado Inserto N) y un Inserto 2 en el dominio catalítico inusualmente largo. Estas dos nuevas propiedades pueden ser las responsables de la preferencia por sustratos largos.
- Xd- α Glu expresada en *S. cerevisiae* presenta las mismas características de estabilidad y actividad que la enzima ortóloga.

La autora de esta tesis ha realizado:

- Caracterización estructural de Xd- α Glu mediante modelado bioinformático por homología de secuencia. Determinación de los elementos estructurales comunes con los miembros GH31, así como los posibles determinantes estructurales responsables de la actividad de Xd- α Glu, específica por sustratos largos.
- Estudio de las secuencias consenso que determinan los residuos catalíticos, así como otros motivos estructurales implicados en la actividad. Alineamiento de las secuencias de distintas α -glucosidasas, y estudio de las características de las enzimas de eucariota que determinan la especificidad por sustratos tipo maltósidos frente a isomaltósidos.

Molecular characterization and heterologous expression of a *Xanthophyllomyces dendrorhous* α -glucosidase with potential for prebiotics production

Patricia Gutiérrez-Alonso¹ · María Gimeno-Pérez¹ · Mercedes Ramírez-Escudero² · Francisco J. Plou³ · Julia Sanz-Aparicio² · María Fernández-Lobato¹

Received: 4 August 2015 / Revised: 9 November 2015 / Accepted: 11 November 2015 / Published online: 28 November 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Basidiomycetous yeast *Xanthophyllomyces dendrorhous* expresses an α -glucosidase with strong transglycosylation activity producing prebiotic sugars such as panose and an unusual tetrasaccharides mixture including α -(1–6) bonds as major products, which makes it of biotechnological interest. Initial analysis pointed to a homodimeric protein of 60 kDa subunit as responsible for this activity. In this study, the gene *Xd-AlphaGlu* was characterized. The 4131-bp-long gene is interrupted by 13 short introns and encodes a protein of 990 amino acids (Xd-AlphaGlu). The N-terminal sequence of the previously detected 60 kDa protein resides in this larger protein at residues 583–602. Functionality of the gene was proved in *Saccharomyces cerevisiae*, which produced a protein of about 130 kDa containing Xd-AlphaGlu sequences. All properties of the heterologously expressed protein, including thermal and pH profiles, activity on different substrates, and ability to produce prebiotic sugars were similar to that of the

α -glucosidase produced in *X. dendrorhous*. No activity was detected in *S. cerevisiae* containing exclusively the 1256-bp from gene *Xd-AlphaGlu* that would encode synthesis of the 60 kDa protein previously detected. Data were compatible with an active monomeric α -glucosidase of 990 amino acids and an inactive hydrolysis product of 60 kDa. Protein Xd-AlphaGlu contained most of the elements characteristic of α -glucosidases included in the glycoside hydrolases family GH31 and its structural model based on the homologous human maltase-glucoamylase was obtained. Remarkably, the Xd-AlphaGlu C-terminal domain presents an unusually long 115-residue insertion that could be involved in this enzyme's activity against long-size substrates such as maltoheptaose and soluble starch.

Keywords *Xanthophyllomyces dendrorhous* · Alpha-glucosidase · GH31 family · Maltooligosaccharides · Panose

Electronic supplementary material The online version of this article (doi:10.1007/s00253-015-7171-3) contains supplementary material, which is available to authorized users.

✉ María Fernández-Lobato
mfemandez@cbm.csic.es

¹ Centro de Biología Molecular Severo Ochoa, Departamento de Biología Molecular (CSIC-UAM), Nicolás Cabrera 1. Universidad Autónoma Madrid, 28049 Madrid, Spain

² Departamento de Cristalografía y Biología Estructural, Instituto de Química-Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain

³ Instituto de Catálisis y Petroleoquímica, CSIC, Marie Curie 2, 28049 Madrid, Spain

Introduction

Glycosidases are widely used as biocatalysts in the biotechnological industries using oligo- and polysaccharides as raw material. α -Glucosidases (EC. 3.2.1.20) are a particularly important subset of these enzymes that hydrolyze α -glycosidic linkages from the non-reducing end of different size substrates. They show major roles in biology ranging from breakdown of polysaccharides to biosynthesis of glycoproteins (Chiba 1997; Melo et al. 2006). According to the Carbohydrate-Active Enzymes database (CAZy) glycoside hydrolase (GH) classification, the α -glucosidases are included into two major families, GH13 and GH31, and, to a lesser extent, in families GH4, GH63, and GH97 (Lovering et al. 2005). Family GH13 includes important enzymes such as α -amylases and

glucosyltransferases that have attracted notable attention during the last few decades, with numerous mechanistic studies and three-dimensional structures solved (Henrissat and Davies 1997; Pal et al. 2010; Janecek et al. 2014). α -Glucosidases from *Geobacillus* sp. strain HTA-462 (Shirai et al. 2008) and *Halomonas* sp. strain H11 (Shen et al. 2014) are included in this family. By contrast, family GH31 includes important enzymes such as the human lysosomal α -glucosidase (whose deficiency results in the Pompe's disease; Hermans et al. 1991; Raben et al. 2002) or the sucrose-isomaltase (target of inhibition of some anti-diabetes drugs; Mohan et al. 2014), much less data concerning this protein family has been clarified.

Although most α -glucosidases exhibit preference for disaccharides and trisaccharides as substrates, a few members, particularly those belonging to GH31 and GH4 families, display specificity for carbohydrates with a high degree of polymerization (Tagami et al. 2015). The α -glucosidase from the basidiomycetous yeast *Xanthophyllomyces dendrorhous* (formerly *Phaffia rhodozyma*) is an extracellular glycoprotein of about 120 and 60 kDa in native and denaturing conditions, respectively, able to hydrolyze maltooligosaccharides and soluble starch (Marín et al. 2006). An outstanding peculiarity of this enzyme is its transglycosylation activity yielding oligosaccharides with α -(1–2), α -(1–4), and α -(1–6) bonds, including prebiotic oligosaccharides such as panose, 6-O- α -glucosyl-maltotriose and 6-O- α -isomaltosyl-maltose as major products (Fernández-Arrojo et al. 2007). This product pattern enriched in panose and unusual tetrasaccharides contrasts with that obtained with most microbial α -glucosidases, which synthesize basically isomaltose with lesser amounts of panose (Wang et al. 2009), and gives biotechnological interest to the *X. dendrorhous* enzyme. To understand the particular behavior of this enzyme, and to improve its properties in the future, a molecular-structural characterization is essential. Here, we have isolated, characterized, and analyzed the gene responsible for this activity; proven its functionality in *Saccharomyces cerevisiae*; and studied the characteristics of the heterologously produced protein. A structural model of this enzyme based on the homologous human maltase-glucoamylase has also been obtained.

Materials and methods

Organisms, transformations, and culture conditions

Xanthophyllomyces dendrorhous ATCC MYA-131 was grown at 23 °C on yeast extract peptone dextrose (YEPD) medium (1 % yeast extract, 2 % peptone, 2 % glucose; all w/v) or maltose minimal medium (MMM) (0.67 % yeast nitrogen base [YNB], 2 % maltose; all w/v). *Saccharomyces cerevisiae* EUROSCARF Y02321 [BY4741; *MATa his3 Δ 1*

leu2 Δ 0 met15 Δ 0 ura3 Δ 0 YIL162w(SUC2)::kanMX4] and BY4742 (*MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0*) were used as expression hosts, and were grown at 30 °C on YEPD. Minimal SC(U)D medium (0.67 % YNB, 2 % glucose, 0.1 % leu, 0.05 % his and 0.05 % met, or 0.1 % lys; all w/v) was used for selection after transformation by lithium acetate method (Burke et al. 2000). To induce protein expression minimal SC(U)Gal and rich YEPGal media were used (same as SC(U)D and YEPD but galactose instead of glucose). Growth was monitored at 660 nm (OD₆₆₀). *Escherichia coli* DH5 α [λ^- ϕ 80d Δ lacZ Δ M15 Δ (*lacZ*-*argF*)*U169 recA1 endA1 hsdR17(r_K⁻ m_K⁻)*] was used for DNA manipulation using standard techniques.

DNA manipulations and cloning of the *X. dendrorhous* α -glucosidase gene

The coding sequence of the α -glucosidase from *X. dendrorhous* was amplified by PCR from a complementary DNA (cDNA) library generated with the pBluescript II XR cDNA library construction kit (Stratagene, Cedar Creek, TX, USA) as previously referred (Linde et al. 2009). Initially, a 140-bp fragment of the gene was amplified using the GlucoSecA(+) and GlucoSecC(–) primers (Table 1) that were directed against two peptides predicted from matrix-assisted laser desorption ionization-time-of-flight-mass spectrometry (MALDI-TOF) analysis. Expand Long Template PCR System (Roche, Indianapolis, IN, USA) was used. The amplified fragment was purified from agarose gel, cloned into pST-Blue1 vector (Novagen, San Diego, CA, USA) and sequenced (SIDI, Universidad Autónoma de Madrid, Madrid, Spain). For subsequent PCRs, the primers GlucoSecC(+) and GlucoSecA(–) (Table 1) combined with the T7 and T3 universal primers (directed towards the vector including cDNA library) were used. All amplified fragments were subjected to the process described earlier resulting in a 2973-bp sequence (an open reading frame of 2970-bp), which was finally amplified from the cDNA library using primers SigPAlpha(+) and AlphaEnd(–). Standard PCR conditions were as follows: (i) 94 °C for 2 min; (ii) 3 cycles of 94 °C for 10 s, the appropriate annealing temperature for each primer pairs for 30 s, 68 °C for 2 min; (iii) 35 cycles, the first as in the previous conditions, which gradually increased by 20 s per cycle at 68 °C; and finally, (iv) 68 °C for 7 min.

To characterize the genomic DNA encoding the putative α -glucosidase, total yeast DNA was extracted and purified from a 16-h-grown culture as described before (Linde et al. 2009) and used as template in PCR reactions. Inverse PCR was used to analyze the flanking sequences of the first 140-bp fragment initially obtained. Briefly, genomic DNA was digested with *EcoRV* (an enzyme that had no restriction site in the 140-bp fragment). Linear DNA (0.2 μ g/mL) was circularized using T4 DNA ligase (20 U/mL; Roche, Mannheim,

Table 1 Primers used in this study

Primer	Sequence
GlucSecA (+)	GCYTTYATYGAYGAYGGWGA
GlucSecA (−)	CWCCRTCTCRATRAARGC
GlucSecC (+)	GCYCARGTGACYGTCYTRGG
GlucSecC (−)	CCYARGACRGTCACYTGRGC
PCR2.1 (+)	ATATAATGTTGGACAGAACTTGCTCAAGTGAC
PCR1.1 (−)	AGAATCGAATAGGGAGATTCTCTGGTATCGTATAC
SigPAlpha (+)	ATGTCGTCCAGATTCAAGACGACC
AlphaEnd (−)	TACGCCCAATCAAGCTCCC
Bam-SigPAlpha (+)	TAATAAGGGATCCAAATGT GTCCAGATTCAAGACG
Py-Xba-Alpha (−)	CTTGGTCTAGATTACGCCCAATCAAGCTCCCAAC
Bam-GluNter (+)	TAATAAGGGATCCAAATGGTCGGTCAAGGCGTGGAT

Restriction sites are underlined. *W*: A or T, *Y*: C or T, *R*: A or G

Germany) and used as template with primers PCR2.1(+) and PCR1.1(−) (Table 1), both were directed to the DNA ends to be characterized. A 5317-bp fragment of DNA was finally analyzed that contained a putative α -glucosidase gene (*Xd-AlphaGlu*) of 4131-bp. All PCR products were introduced into the pST-Blue1 vector and sequenced as referred.

To express the potential α -glucosidase gene in a heterologous system, a fragment of about 3000-bp was amplified from the cDNA library using primers Bam-SigPAlpha(+) and Py-Xba-Alpha(−) (Table 1). The amplification product, flanked by *Bam*HI and *Xba*I sites, was digested with the indicated enzymes and included in the pYES2 plasmid (Invitrogen, Carlsbad, CA, USA) under *pGAL1* promoter control (construction Xd-AG-pYES). Primers Bam-GluNter(+) and Py-Xba-Alpha(−) were used to amplified a 1256-pb fragment from the mentioned library, which would result in the last 408 residues of the potential protein. The fragment was included in pYES2 as discussed earlier (construction Xd-AGNter-pYES). Constructions were verified by sequencing and used to transform the *S. cerevisiae* strains BY4741 and BY4742. Transformants including empty pYES2 plasmid were also obtained and used as controls.

Protein purification, analysis and quantification

α -Glucosidase from *X. dendrorhous* was purified as previously described (Marín et al. 2006). Basically, yeasts were grown in MMM medium (1 L; 4 U/mL; 2 U/ μ g), the extracellular fraction was concentrated (about five times; 17 U/mL; 3 U/ μ g) through a 30000 molecular weight cutoff (MWCO) PES membrane in a Vivaflow 50 system (Sartorius, Goettingen, Germany), dialyzed in 20 mM HCl-Tris pH 7 (buffer A), and applied to a DEAE-Sephacel chromatography column (10 mL) equilibrated with buffer A. Proteins were eluted with a discontinuous

gradient of 0–0.2 M NaCl. Active fractions eluting at 0.05 M salt were dialyzed in 20 mM sodium acetate pH 4.5 (buffer B), applied to the DEAE-Sephacel equilibrate in this buffer and eluted at 0.2 M salts (15 U/mL; 106 U/ μ g). The protein profiles were determined measuring absorbance of fractions at 280 nm (NanoDrop spectrophotometer ND-1000, Thermo Fisher, Wilmington, DE, USA). To analyze the α -glucosidase activity from *X. dendrorhous* expressed in *S. cerevisiae*, host yeast was grown in 100 mL of SC(U)D medium to ~ 1.0 OD₆₆₀ and protein expression was induced in YEPGal at 30 °C during 26 h. Cells were removed at 6000 \times g for 15 min, and extracellular fraction was concentrated (about 100-fold) using a Vivaflow 50 system as referred earlier followed, if required, by MWCO 100 kDa Microcon system (Millipore, Cork, Ireland; concentrated 10 times more). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; 8 % polyacrylamide) and ProtoBlue Safe Colloidal Coomassie stain (National Scientific, Atlanta, GA, USA) were used to analyze the purity of the proteins. Precision Plus Protein standards 10–250 kDa or SDS-PAGE Standards broad range 6.5–200 kDa (Bio-Rad, Hercules, CA, USA) were used as molecular weight standards. Protein concentration was determined using the Bio-Rad microprotein determination assay according to the manufacturer's specifications and bovine serum albumin as a standard. Proteins were excised from SDS-PAGE gels, digested with trypsin, and analyzed by matrix-assisted laser desorption ionization-time-of-flight-mass spectrometry (MALDI-TOF; Autoflex, Bruker, Bremen, Germany) at the Proteomic Service of “Centro de Biología Molecular Severo Ochoa” (Madrid, Spain). The tryptic peptide map obtained was assigned by comparing their masses with those calculated from theoretical tryptic digestion of protein α -glucosidase. Assignment was verified by analyzing the peptides by reverse-phase LC coupled to MS (RP-LC/MS) using a Deca XP mass spectrometer (Thermo-Finishing, San José, CA, USA) and a ThermoHypersil (0.18 \times 150 mm) C18 column (Thermo Fisher, Wilmington, DE, USA). The mass spectrometer was operated in the selected MS/MS ion monitoring mode and the spectra from the peptides were analyzed by assigning the fragments to the candidate sequence after calculating the series of theoretical fragmentations. For the N-terminal amino acid sequencing, the proteins (about 500 ng) were subjected to SDS-PAGE and blotted onto polyvinylidene difluoride membranes (Millipore). Membranes were stained with Coomassie brilliant blue R250 and the band obtained was excised and sequenced (Helmholtz Center for Infection Research, Braunschweig, Germany).

Enzyme activity assay

Unless otherwise indicated, α -glucosidase hydrolytic activity was determined by measuring the amount of glucose liberated from maltose 1 % (w/v) in 50 mM sodium phosphate pH 5.5 at 45 °C for 90 min. The mixture was boiled for 10 min at 100 °C,

and glucose was quantified by using the glucose oxidase-peroxidase assay under the conditions described by Marín et al. (2006). One unit of α -glucosidase activity was defined as that corresponding to the release of 1 μ mol of glucose per minute under conditions described earlier. Maltotriose, maltoheptaose, maltodextrin (Sigma, St. Louis, MO, USA), and soluble starch (Difco; Detroit, MI, USA) were also used as substrates in similar reactions. Estimation of α -glucosidase activity at different pH values (3.0–9.0) and temperatures (20–80 °C) was carried out under the aforementioned conditions using maltose as substrate and 45 °C and pH 5.5, respectively. Buffers used were glycine (pH 3.0–3.5), citric acid-sodium citrate (pH 3.5–5.0), Na_2HPO_4 - NaH_2PO_4 (pH 5.0–7.5) and Tris-HCl (pH 7.5–8.0), all at 50 mM. The thermostability refers to the temperature required for 50 % enzyme inactivation (enzyme half-life) after heating the pure enzyme at different temperatures (35–80 °C) during 30–120 min, and was determined removing samples at regular intervals and estimating the residual α -glucosidase hydrolytic activity. Intracellular activity was evaluated after cell lysate using Yeast Buster™ (Novagen, San Diego, CA, USA) according to the manufacturer's protocol.

Production of isomaltooligosaccharides

Transferase activity was determined using maltose 200 g/L in 0.1 M sodium acetate pH 5.6. Reaction volume was 2 mL, and final activity in mixture was 20–30 mU/mL. Mixtures were incubated at 45 °C and 100 rpm on an orbital shaker (Vortemp 56, Labnet International, Edison, NJ, USA). Aliquots (40 μ L) were withdrawn at different times and enzyme was inactivated for 10 min at 100 °C. Samples were diluted with water (1:5, 1:20, and 1:100) to a final volume of 200 μ L, centrifuged for 5 min at 2400 \times g through a 0.45 μ m Durapore® membrane (Millipore, Darmstadt, Germany), and analyzed by HPLC with a quaternary pump (Delta 600, Waters, Cerdanyola del Vallès, Barcelona, Spain) coupled to a 5 μ m Lichrosorb-NH₂ column (4.6 \times 250 mm; Merck, Darmstadt, Germany as indicated (Fernández-Arrojo et al. 2007). Maltose, glucose, maltotriose, maltotetraose, panose, kojibiose, and isomaltose (Sigma, St. Louis, MO, USA) were used as standards.

Phylogenetic analysis and molecular modeling

Sequence encoding the α -glucosidase from *X. dendrorhous* have been assigned the GenBank accession no. KT223785. Amino acid sequence of α -glucosidase was blasted against protein database at Swiss-Prot (<http://www.expasy.org/tools/blast/>) and sequences were aligned with CLUSTALW interface in MEGA4.0 [<http://www.megasoftware.net/>] (pair wise alignment gap opening penalty 10; gap extension penalty 0.1; multiple alignment gap opening penalty 10; gap extension penalty 0.2). Bootstrap test of phylogeny was used

with the tree obtained. A structural model of *X. dendrorhous* α -glucosidase, comprising residues 64–990, was performed with the SWISS-MODEL server (Arnold et al. 2006). Best template was the N-terminal domain of human intestinal maltase-glucoamylase (PDB code 2QLY; Sim et al. 2008).

Results

Isolation, characterization and analysis of the *Xd-AlphaGlu* gene from *X. dendrorhous*

To isolate the gene encoding the α -glucosidase activity from *X. dendrorhous*, the enzyme was purified as indicated in “Materials and methods.” Interestingly, and differing from what we reported earlier (Marín et al. 2006), a protein of about 120 kDa was clearly visualized on SDS-PAGE in addition to the expected one of 60 kDa (Fig. 1). The two bands were processed for amino acid sequencing by tryptic digestion followed by MALDI-TOF-MS analyses (data not shown). The protein of 60 kDa retrieved two peptides with the predicted sequences LAQVTVLGVK (1027.6 m/z) and SSEAEGYAFIDDGESFPVTDSR (2379 m/z), which aligned with part of the amino acid sequences of putative α -glucosidases from fungi as *Chaetomium thermophilum* (XP_006695829.1; query cover 100 %, identity 90 %) and *Phanerochaete carnosae* (XP_007389717.1; query cover 94 %, identity 57 %). Both sequences were also localized in the *Aspergillus niger* enzyme (P56526; identities 89 and 62 % for the 1027.6 and 2379 m/z peptides, respectively). The protein of 120 kDa generated the peptide VIEQYSEISGR (1280.6 m/z), which aligned with sequences of α -glucosidases from plants as *Populus euphratica* (XP_011010437.1, query cover 100 %, identity 82 %). In addition, the N-terminus of the 60 kDa protein was

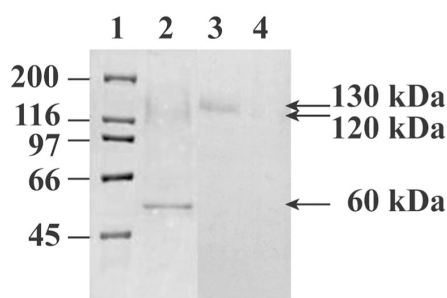


Fig. 1 SDS-PAGE analysis of the α -glucosidase protein. Enzyme purified (0.5 μ g) from *X. dendrorhous* (lane 2), extracellular fraction of *S. cerevisiae* including construction *Xd-AG-pYES* (lane 3) and *pYES2* (lane 4) were concentrated about 100 times as indicated in the “Materials and methods” section. Numbers at the left indicate the positions of molecular mass standards (lane 1) in kilodaltons. Masses of the proteins expressed in the two yeasts are indicated at the right

determined to be VGQGV DVNTPAYEIHNALDK, which also aligned with a glycoside hydrolase from the fungus *Trametes cinnabarina* (CDO74837.1; query cover 65 %; identity 77 %) whereas no data were obtained from the N-terminus of the 120 kDa protein.

The primers GlucoSecA(+) and GlucoSecC(−) (Table 1) both directed to part of the sequences obtained from the 60 kDa protein analysis (AFIDDGE and AQVTVLG) were used to amplify a 140-bp DNA fragment from a *X. dendrorhous* cDNA library, which encoded a sequence that showed similarities with internal fragments of α -glucosidases from filamentous fungi and yeasts such as *Postia placenta* (XP_002469070.1; query cover 98 %, identity 48 %) and *Saccharomycopsis fibuligera* (CAF31354.1; query cover 100 %, identity 42 %). By using the strategy described in the “Materials and methods” section, an open reading frame (ORF) of 2970-bp was identified (*Xd-AlphaGlu*) from the cDNA library, corresponding to a 990 amino acid polypeptide. This cDNA sequence was also localized in a fragment of 4131-bp of genomic DNA. Comparing sequences of both fragments, 13 introns were identified, 12 of them located in the first half of the genomic interrupted sequence (Fig. 2), with an average size of about 100-bp and canonical splicing sites (Table S1 in the Supplementary Material). The total sequence from genomic DNA characterized in this work was 5317-bp, of which 686-bp preceded the characterized ORF (Fig. 2), 4131-bp encoded the *Xd-AlphaGlu* protein and 500-bp were located downstream below the stop codon TAA. Two putative carbon source-responsive elements (CSRE) were located in the position −618/−603 and −303/−288 (Cartharius et al. 2005), which probably are involved in the catabolic repression of *Xd-AlphaGlu* gene. Indeed, no α -glucosidase activity was previously detected in *X. dendrorhous* cultures by using glucose as carbon source (Marin et al. 2006).

Analysis of the deduced amino acid sequence encoded by the *Xd-AlphaGlu* gene

The 990 amino acid deduced polypeptide characterized here has a predicted molecular mass of 110.4 kDa, an isoelectric point of 4.64 units, includes a possible signal peptide for protein export of 31 amino acids (using SignalP 3.0. <http://www.cbs.dtu.dk/services/SignalP/>) and 19 potential N-glycosylation sites within the consensus sequence Asn-Xaa-Ser/Thr, where Xaa is any amino acid other than proline (Fig. 2). In addition, all predicted sequences earlier referred for the 60 kDa protein, the two peptides resulting from the MALDI-TOF analyses and the N-terminal sequence were located in the 990 amino acid polypeptide (residues 894–912,

937–946, and 583–602, respectively), as well as the peptide proposed for the 120 kDa protein (residues 328–338). Positioning of peptide masses generated by the proteins MALDI-TOF analyses are showed as supplemental material (Fig. S1 in the Supplementary Material).

The overall deduced protein sequence of *Xd-AlphaGlu* was very similar to that of the α -glucosidases from ascomycetous yeasts and *Aspergillus* spp. Indeed, it was most similar to enzymes from *Candida tsukubaensis* (40 % identity over a 970 amino acid overlap), *A. oryzae* (38 % identity over a 921 amino acid overlap), *A. niger* (38 % identity over a 921 amino acid overlap), *C. albicans* (37 % identity over a 960 amino acid overlap) and other yeasts such as *Schwanniomyces occidentalis* (39 % identity over a 910 amino acid overlap) or *Schizosaccharomyces pombe* (39 % identity over a 910 amino acid overlap). Lower similarities were also observed to proteins from plants and mammals (Fig. 3). Closest structurally known homologues are the N-terminal domains of human maltase-glucoamylase (hNt-MGAM) and the sucrose-isomaltase (hNt-SI), both sharing 35 % identity over an 891 amino acid overlap. A model was built using hNt-MGAM as a template. According to it, *Xd-AlphaGlu* contains most of the elements characteristic of α -glucosidases included in the glycoside hydrolases family GH31 (Fig. 4), these being a major catalytic (β/α)₈ barrel domain and the N- and C-terminal domains, with one or two β -sandwich structures, respectively. However, two distinct structural features were found. First, the N-terminal domain contains a 30 residues insertion in a loop linking two β -sheets (insert N in Fig. 4) that bulges from the domain. Second, and in common with GH31 members from eukarya (Sim et al. 2008; Tagami et al. 2013), the catalytic (β/α)₈ barrel domain has two inserted regions. Insert 1 is located at its loop β 3– α 3, being equivalent to that in hNt-MGAM. Insert 2 is placed at loop β 4– α 4 being highly variable in shape and length within the GH31 family. Remarkably, *Xd-AlphaGlu* presents an unusually long 115 residues segment at Insert 2, which is probably spatially close to the N-insert described earlier.

Functional analysis of the *X. dendrorhous Xd-AlphaGlu* gene

The functionality of the *Xd-AlphaGlu* gene was analyzed in two *S. cerevisiae* strains lacking the YGR292W gene responsible for the α -glucosidase MAL12, both unable to show α -glucosidase activity on a galactose-based medium. The analyzed sequence from *X. dendrorhous* encoding for a putative 990 amino acid α -glucosidase, from the initial Met₁ (1ATG₃) to Ala₉₉₀

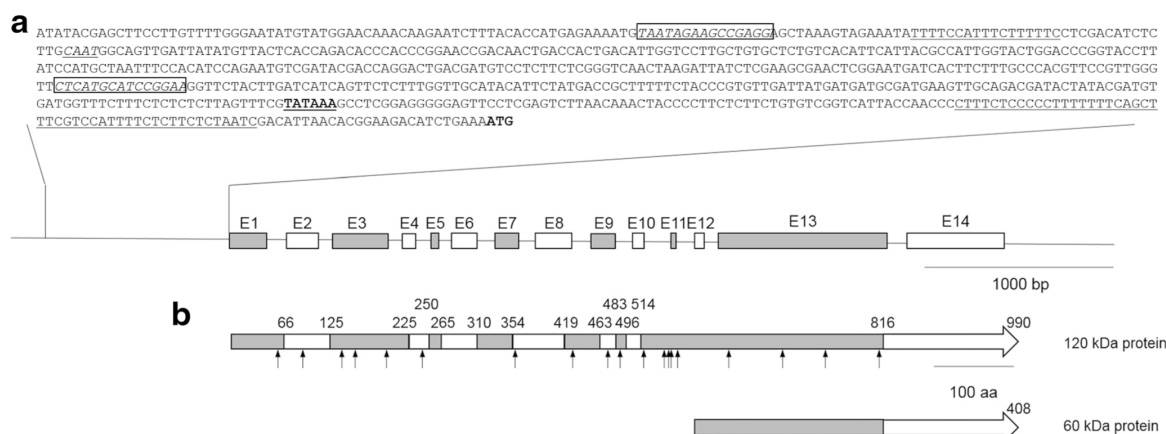


Fig. 2 Scheme of DNA and amino acid sequence of the α -glucosidase from *X. dendrorhous*. **a** Map of the genomic DNA region of *X. dendrorhous* including the *Xd-AlphaGlu* gene. Exons encoding the deduced protein are shown as boxes from *E1* to *E14*. Potential regulatory sequence identified in the 5' non-coding region are indicated as follows: TATA-like box in **bold** and underlined, CAAT boxes in *italics* and underlined, pyrimidine-rich stretches underlined, carbon source-

responsive elements in *italics* and **boxed**. ATG translation start codon is shown in **bold**. **b** Schematic representation of the 120 and 60 kDa *Xd-AlphaGlu* protein. Amino acid sequence codified for different exons is represented as *white* or *gray* boxes. Arrows pointed to the putative N-glycosylation sites are indicated (positions 61, 97, 144, 165, 199, 245, 358, 431, 474, 486, 519, 545, 554, 558, 564, 611, 671, 750, and 811)

(²⁹⁶⁸GCG²⁹⁷⁰), was included in the pYES2 vector under *pGAL1* promoter control (construction *Xd-AG-pYES*). α -Glucosidase activity was detected in positive transformants (3–5 from each strain), but only weak

activity (4–10 mU/mL) was quantified in the yeast's extracellular and cellular fractions. Figure 5 shows data obtained with one of the selected transformants. Maxima levels of activity were detected in the culture filtrates at the

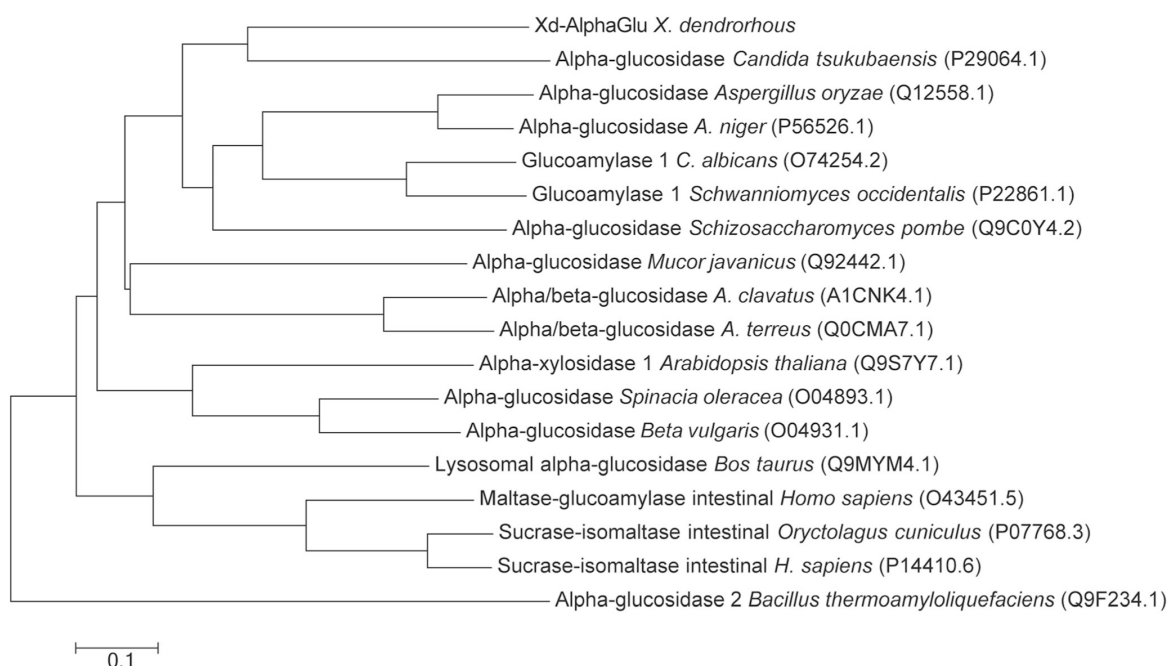
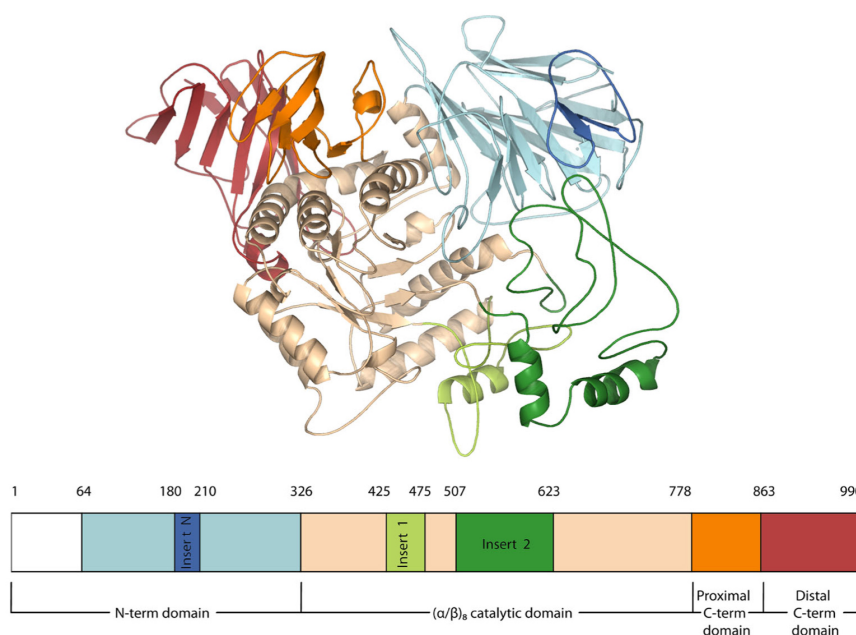


Fig. 3 Phylogenetic analysis of the α -glucosidase from *X. dendrorhous*. The radical tree was constructed from the alignment of the amino acid sequences of proteins with proven α -glucosidase activity. The GenBank accession numbers are indicated

Fig. 4 Overall structure of Xd-AlphaGlu. The model, built using the hNt-MGAM as a template (*top image*), keeps the color code given in the sequence scheme (*lower image*). A 30-residue insert at the N-terminal domain, and two inserts found at the catalytic domain of GH31 from eukarya, are highlighted in *blue* and *green*. Most outstanding features of Xd-AlphaGlu are the long segments named N-insert and Insert 2



beginning of the culture stationary phase (7.5–8.5 OD₆₆₀), where basically only one protein of about 130 kDa was visualized on SDS-PAGE (Fig. 1). Analysis by MALDI-TOF and fingerprinting of this protein retrieved peptides that were located in the Xd-AlphaGlu sequence, such as the NASDDETFIDTRPSSK (1879.9 m/z) and the RPFIVSR (874.5 m/z) (Fig. S1 in the Supplementary Material). No α -glucosidase activity was detected in the transformants including construction Xd-AGNter-pYES, which theoretically would produce the protein of 60 kDa previously detected in *X. dendrorhous* (from V₅₈₃ to Ala₉₉₀ of Xd-AlphaGlu), and that started with the sequence initially characterized by N-terminal sequencing. As expected, no intra or extracellular activity was detected in any of the yeasts including the empty pYES2 vector.

Analysis of the α -glucosidase activity expressed in *Saccharomyces cerevisiae*

Biochemical characteristics of the α -glucosidase activity expressed in the heterologous system were compared with those shown by the protein expressed in *X. dendrorhous*. Yeasts were grown and proteins were obtained from the extracellular fractions as described in the “Material and methods” section. Regardless, the producing yeast (*S. cerevisiae* or *X. dendrorhous*), the enzyme displayed maximum activity at pH 5.0–5.8 and 40–48 °C. In addition, the enzymes from the two yeasts showed a similar profile of thermostability in the range

of 35–70 °C (without the maltose substrate), with a 50 % activity (T₅₀) of 40–80–120 min in the 65–55–45 °C range and no activity at 70 °C (data not shown). Hydrolytic activity of the proteins produced by the two yeasts was also evaluated using different size substrates (Table 2) and both were similar, showing preference for maltose and the larger substrates analyzed, maltoheptaose and soluble starch.

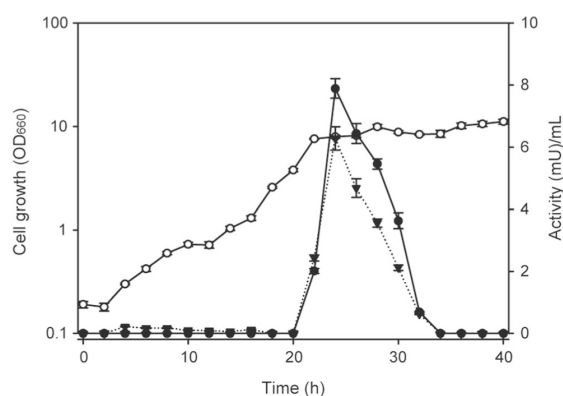


Fig. 5 Time course of α -glucosidase activity detected in *S. cerevisiae*. Strain BY4742 was transformed with Xd-AG-pYES construction and cultivated in 100 mL of YEPGal (*empty circles*). Samples were withdrawn at the indicated times and analyzed for cellular (*filled triangles*) and extracellular (*filled circles*) α -glucosidase activity using maltose as substrate. Each point represents the average of three independent measurements; standard deviation is indicated

Table 2 Xd-AlphaGlu activity detected in yeasts

Substrate	α -Glucosidase activity (%)	
	<i>S. cerevisiae</i>	<i>X. dendrorhous</i>
Maltose	100.0 \pm 3.5	100.0 \pm 2.0
Maltotriose	59.9 \pm 1.3	55.1 \pm 4.5
Maltoheptaose	82.5 \pm 6.7	85.3 \pm 4.6
Maltodextrin	67.5 \pm 5.2	64.1 \pm 5.2
Soluble starch	82.4 \pm 7.0	80.8 \pm 5.8

Value are the average of at least three measures \pm standard deviation. One hundred percent activity represents 7 mU/mL

The glucosyltransferase capacity of the enzyme expressed in *S. cerevisiae* was evaluated using 200 g/L maltose, and isomaltooligosaccharides production was observed from the beginning of the reaction. At 24 h, 6.4 and 4.2 g/L of maltotriose and panose were estimated, respectively. Amounts gradually increased for 7 days reaching 23.7 and 25.9 g/L, respectively, which indicates that the enzyme is still active in the used condition. In the reaction mixture, tetrasaccharides such as 6-O- α -glucosyl-maltotriose (6.0 g/L) and 6-O- α -isomaltosyl-maltose (7.9 g/L) were also quantified (Fig. 6). The profile and amounts of the oligosaccharides synthesized by the heterologously expressed enzyme were also similar to those obtained by the enzyme expressed in the natural producer (Fernández-Arrojo et al. 2007).

Discussion

Here, we have isolated and characterized the gene *Xd-AlphaGlu*, responsible for the α -glucosidase activity from *X. dendrorhous*. The sequence described showed 13 introns with an average size within the range of introns described in fungi and yeasts (100–120-bp;

Bon et al. 2003). The intron rate represented about 28 % of the sequence analyzed. The data was not surprising since presence of multiple introns in *X. dendrorhous* genes is a quite common event (Baeza et al. 2015; Loto et al. 2012). The vast majority of the intron splice sites described in this yeast displays the canonical dinucleotide intron borders 5 GT/AG-3, and the branch site consensus sequence TRAY, where A is the branch point which generally is present in the fungal introns (Kupfer et al. 2004). These consensus sequences were also localized in all the *Xd-AlphaGlu* introns (Table S1 in the Supplementary Material). Functionality of the sequence here characterized was clearly proved in *S. cerevisiae* despite the α -glucosidase activity detected in the heterologous system was really small, only 9 mU/mL versus the near 4 U/mL obtained in *X. dendrorhous* (Marín et al. 2006). The difference may be caused by such diverse factors as improper maturation, glycosylation pattern, or even possible degradation of the heterologous protein in the host yeast, among others (Kofod et al. 1994; Rehm et al. 1998; Bang et al. 1999; Rosenberg et al. 1993).

The sequence of the protein Xd-AlphaGlu contained all the peptides previously predicted for the proteins of 120 and of 60 kDa detected in *X. dendrorhous* extracellular media (Fig. S1 in the Supplemental Material), as well as the N-terminal sequence of the smaller one. Our data would be compatible with an active monomeric α -glucosidase of about 120 kDa, and an inactive peptide of 60 kDa, which might result by hydrolysis of the active protein. The extracellular aspartic protease already described in *X. dendrorhous* (Bang et al. 1999) could be responsible for the partial hydrolysis of the enzyme obtained from the natural producer.

Biochemical and enzymatic characteristics of the activity expressed in *S. cerevisiae* were virtually identical to those of the produced in *X. dendrorhous*, including its glucosyltransferase ability. Glucooligosaccharides obtained in the transglycosylation reaction containing α -(1–6) linkages (isomaltooligosaccharides, IMOS) that present some advantages over other established prebiotics. In particular, they are more stable at different pH values and contribute to a lower gas release (Goffin et al. 2011). The profile of the products obtained in the transglycosylation reaction with the heterologous enzyme is also a clear proof that the gene *Xd-AlphaGlu* characterized here is responsible for the α -glucosidase activity previously detected in *X. dendrorhous* (Marín et al. 2006; Fernández-Arrojo et al. 2007). Indeed, even if the used host yeast displayed any trace of α -glucosidase activity (and this is not the case), the enzyme from *S. cerevisiae* produces a very different pattern of IMOS, where isomaltose is the only

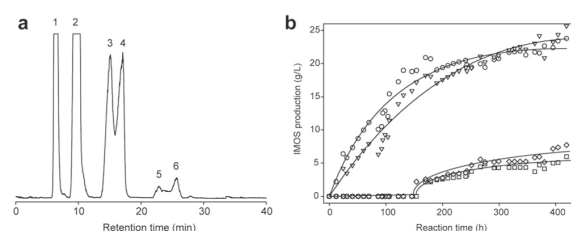


Fig. 6 Analysis of the transglucosylation reaction mediated by the Xd-AlphaGlu expressed in *S. cerevisiae*. **a** HPLC chromatogram of reaction mixture after 419 h. Peaks: 1, glucose; 2, maltose; 3, maltotriose; 4, panose; 5, 6-O- α -glucosyl-maltotriose; 6, 6-O- α -isomaltosyl-maltose. **b** Time course of oligosaccharide production: panose (triangles), maltotriose (circles), 6-O- α -glucosyl-maltotriose (squares), 6-O- α -isomaltosyl-maltose (rhombus)

significant product obtained (Fernández-Arrojo et al. 2007). Commercially available IMOS are mainly composed of isomaltose, panose and isomaltotriose (Pan and Lee 2005). Interestingly, the sugar mixture obtained with the *X. dendrorhous* enzyme is composed of tri- and tetrasaccharides containing α -(1–4) and α -(1–6) linkages, while other α -glucosidases basically synthesize disaccharides and minor amounts of trisaccharides (Kato et al. 2002), some of them majorly isomaltose with variable amounts of panose (Wang et al. 2009). Among the sugars produced by the Xd-AlphaGlu, it is worth mentioning panose, which constitutes about 40 % of the products synthesized. This mildly sweet trisaccharide is considered to be widely used in the food industry as an anti-fading agent due to its anti-cariogenic properties. Panose was also the main sugar obtained by the *A. oryzae* α -glucosidase but no larger size sugar was detected in the reaction mixture (Wu et al. 2010). To our knowledge, no enzyme producing the tetrasaccharides mixture generated by the enzyme from *X. dendrorhous* in a transglycosylation reaction was described, which makes it of biotechnological interest.

The protein encoded by the *Xd-AlphaGlu* gene belongs to the family 31 of the glycoside hydrolases (GH31). Two aspartic residues act as catalytic agents in the hydrolytic reaction mediated by these enzymes, the first as nucleophile and the second as acid-base catalyst. In the Xd-AlphaGlu protein, Asp501 and Asp672 would be these residues, respectively. The region around the residue acting as a nucleophile:

[GFY]-[LIVMF]-WxDM-[NSA]-E is highly conserved in the GH31 family (consensus motifs PS00129 and PS00707, PROSITE) and was also located between positions 495 and 503 of the protein here analyzed. Recently, the genome of *X. dendrorhous* strain CBS6938 (ATCC96594) has been assembled (EMBL: LN483084-LN483350) and showed the presence of at least four sequences related with the GH31 family. One of them, which would result in the amino acid sequence CED83529.1, exhibits high homology with Xd-AlphaGlu (99.8 % identity; Sharma et al. 2015). Our data indicate that this sequence could be responsible for the potential α -glucosidase activity of the analyzed strain.

Reported structural studies showed that GH31 enzymes present a pocket-shaped active site formed mainly by loops linking the β -strands to the α -helices of the catalytic domain, and also, a long loop that bulges from the N-terminal domain (called N-loop). From these, the loop linking β 1– α 1 of the catalytic domain has been related to the ability of some α -glucosidases to hydrolyze α -(1–6) linked substrates, which depends on the presence of a Trp instead of a Tyr. Therefore, this loop has been considered as a substrate specificity region (Tagami et al. 2013). Accordingly, the presence of Trp355 at this position in the Xd-AlphaGlu sequence (Fig. 7a) might explain its ability to produce IMOS showing α -(1–6) linkages. On the other hand, some α -glucosidases are known to display long-chain substrate specificity. This is the case of the α -glucosidase from sugar beet (SBG), and the C-terminal domain of human maltase-glucoamylase (hCt-MGAM). However, different structural determinants governing the long-chain specificity have been described in each case and, thus, residues from the N-loop (SBG) or a longer Insert 1 (hCt-MGAM) have been identified as responsible for shaping subsites +2 and +3 (Tagami et al. 2013; Sim et al. 2008; Ren et al. 2011). Sequence analysis reveals that the key residues responsible for SBG long-chain specificity are missing in the Xd-AlphaGlu N-loop (Fig. 7b) and also that it does not contain the long Insert 1 equivalent to that found in hCt-MGAM. Nevertheless, an outstanding structural feature of Xd-AlphaGlu is the unique long insertion included in Insert 2 (Fig. 4), which might well conform additional binding subsites, explaining its high activity against long substrates as maltoheptaose and soluble starch.

Further research into Xd-AlphaGlu structure-specificity relationships will help in understanding the peculiar activity of this enzyme, and also, it will contribute to shed light on the determinants responsible for glucosyltransferase activity within the GH31 family enzymes.

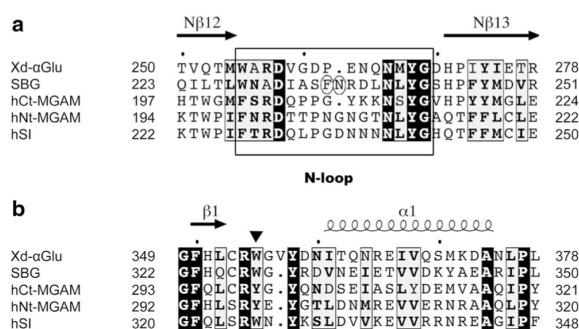


Fig. 7 Alignment of Xd-AlphaGlu regions relevant for specificity. **a** Substrate specificity region. The Trp that has been related to the ability of some α -glucosidases to hydrolyze α -(1–6) linked substrates is marked with a triangle. **b** N-loop region. The residues involved in substrate recognition in SBG are circled. The sequences aligned are as follows: SBG sugar beet α -glucosidase (PDB code 3WEL), hCt-MGAM, C-terminal subunit of human maltase-glucoamylase (PDB code 3TON), hNt-MGAM, N-terminal domains of human maltase-glucoamylase (PDB code 2QLY), hNt-SI, N-terminal domain of human sucrose-isomaltase (PDB code 3LPO). The marked α -helix and β -strand regions correspond to the Xd-AlphaGlu model

Acknowledgments Projects BIO2013-48779-C4-1/-3/-4 from the Spanish Ministry of Economy and Competitiveness supported this research. We thank Fundación Ramón Areces for the institutional grant to the Centro de Biología Molecular Severo Ochoa. M.G.P. was supported by a Spanish FPU fellowship from the Ministry of Economy and Competitiveness.

Compliance with ethical standards

Funding This study was funded by the Spanish Ministry of Economy and Competitiveness (BIO2013-48779-C4-1/-3/-4).

Conflict of interests The authors declare that they have no competing interests.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Arnold K, Bordoli L, Koop J, Schwede T (2006) The Swiss-model workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* 22:195–201. doi:10.1093/bioinformatics/bti770
- Baeza M, Alcaíno J, Barahona S, Sepúlveda D, Cifuentes V (2015) Codon usage and codon context bias in *Xanthophyllomyces dendrorhous*. *BMC Genomics* 16:293. doi:10.1186/s12864-015-1493-5
- Bang ML, Villadsen I, Sandal T (1999) Cloning and characterization of an endo- β -1,3(4)glucanase and an aspartic protease from *Phaffia rhodozyma* CBS 6938. *Appl Microbiol Biotechnol* 51:215–222. doi:10.1007/s002530051384
- Bon E, Casaregola S, Blandin G, Llorente B, Neuveglise C, Münsterkötter M, Güldener U, Mewes HW, Van Helden J, Dujon B, Gaillardin C (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res* 31:1121–1135. doi:10.1093/nar/gkg213
- Burke D, Dawson D, Stearns T (2000) Methods in yeast genetics: A Cold Spring Harbour Laboratory course manual. Cold Spring Harbor Laboratory Press, Plainview, N.Y
- Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21:2933–42. doi:10.1093/bioinformatics/bti473
- Chiba S (1997) Molecular mechanism in α -glucosidase and glucoamylase. *Biosci Biotechnol Biochem* 61:1233–1239. doi.org/10.1271/bbb.61.1233
- Fernández-Arrojo L, Marín D, Gómez de Segura A, Linde D, Alcalde M, Gutiérrez-Alonso P, Ghazi I, Plou FJ, Fernández-Lobato M, Ballesteros A (2007) Transformation of maltose into prebiotic isomaltooligosaccharides by a novel α -glucosidase from *Xanthophyllomyces dendrorhous*. *Process Biochem* 42:1530–1536. doi:10.1016/j.procbio.2007.08.007
- Goffin D, Delzenne N, Blecker C, Hanon E, Deroanne C, Paquot M (2011) Will isomalto-oligosaccharides, a well-established functional food in Asia, break through the European and American market? The status of knowledge on these prebiotics. *Crit Rev Food Sci and Nutr* 51:394–409. doi:10.1080/10408391003628955
- Henrissat B, Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 7:637–644.
- Hermans MM, Kroos MA, van Beeumen J, Oostra BA, Reuser AJ (1991) Human lysosomal α -glucosidase. Characterization of the catalytic site. *J Biol Chem* 266:13507–13512
- Janecek S, Svensson B, MacGregor EA (2014) α -Amylase: an enzyme specificity found in various families of glycoside hydrolases. *Cell Mol Life Sci* 71:1149–1170. doi:10.1007/s00018-013-1388-z
- Kato N, Suyama S, Shirokane M, Kato M, Kobayashi T, Tsukagoshi N (2002) Novel α -glucosidase from *Aspergillus nidulans* with strong transglycosylation activity. *Appl Environ Microbiol* 68:1250–1256. doi:10.1128/AEM.68.3.1250-1256.2002
- Kofod LV, Kauppinen S, Christgau S, Andersen LN, Heldt-Hansen HP, Dörreich K, Dalbøge H (1994) Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*. *J Biol Chem* 269:29182–29189
- Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW (2004) Introns and splicing elements of five diverse fungi. *Eukaryot Cell* 3:1088–1100. doi:10.1128/EC.3.5.1088-1100
- Linde D, Macías I, Fernández-Arrojo L, Plou FJ, Jiménez A, Fernández-Lobato M (2009) Molecular and biochemical characterization of a β -fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol* 75:1065–1073. doi:10.1128/AEM.02061-08
- Loto I, Gutiérrez MS, Barahona S, Sepúlveda D, Martínez-Moya P, Baeza M, Cifuentes V, Alcaíno J (2012) Enhancement of carotenoid production by disrupting the C22-sterol desaturase gene (CYP61) in *Xanthophyllomyces dendrorhous*. *BMC Microbiol* 12:235. doi:10.1186/1471-2180-12-2351471-2180-12-235
- Lovering AL, Lee SS, Kim YW, Withers SG, Strynadka NC (2005) Mechanistic and structural analysis of a family 31 α -glucosidase and its glycosyl-enzyme intermediate. *J Biol Chem* 280:2105–2115. doi:10.1074/jbc.M410468200
- Marín D, Linde D, Fernández Lobato M (2006) Purification and biochemical characterization of an α -glucosidase from *Xanthophyllomyces dendrorhous*. *Yeast* 23:117–125. doi:10.1002/yea.1345S0378-1119(11)00511-7
- Melo EB, Gomes AS, Carvalho I (2006) α - and β -Glucosidase inhibitors: chemical structure and biological activity. *Tetrahedron* 62:10277–10302. doi:10.1016/j.tet.2006.08.055
- Mohan S, Eskandari R, Pinto BM (2014) Naturally occurring sulfonium-ion glucosidase inhibitors and their derivatives: a promising class of potential antidiabetic agents. *Acc Chem Res* 47:211–225. doi:10.1021/ar400132g
- Pal K, Kumar S, Sharma S, Garg SK, Alam MS, Xu HE, Agrawal P, Swaminathan K (2010) Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv glycogen branching enzyme: insights of N-terminal β -sandwich in substrate specificity and enzymatic activity. *J Biol Chem* 285:20897–20903. doi:10.1074/jbc.M110.121707M110.121707
- Pan YC, Lee WC (2005) Production of high-purity isomalto-oligosaccharides syrup by the enzymatic conversion of transglucosidase and fermentation of yeast cells. *Biotechnol Bioeng* 89:797–804. doi:10.1002/bit.20402
- Raben N, Plotz P, Byrne BJ (2002) Acid α -glucosidase deficiency (glycogenosis type II, Pompe disease). *Curr Mol Med* 2:145–166. doi:10.2174/1566524024605789#sthash.KQI1znTD.dpuf
- Rehm J, Willmitzer L, Heyer AG (1998) Production of l-kestose in transgenic yeast expressing a fructosyltransferase from *Aspergillus foetidus*. *J Bacteriol* 180:1305–1310
- Ren LM, Qin XH, Cao XF, Wang LL, Bai F, Bai G, Shen Y (2011) Structural insight into substrate specificity of human intestinal maltase-glucoamylase. *Protein Cell* 2:827–836. doi:10.1007/s12328-011-1105-3
- Rosenberg AH, Goldman E, Dunn JJ, Studier FW, Zubay G (1993) Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *J Bacteriol* 175:716–722

- Sharma R, Gassel S, Steiger S, Xia X, Bauer R, Sandmann G, Thines M (2015) The genome of the basal agaricomycete *Xanthophyllomyces dendrorhous* provides insights into the organization of its acetyl-CoA derived pathways and the evolution of Agaricomycotina. BMC Genomics 16:233. doi:[10.1186/s12864-015-1380-0](https://doi.org/10.1186/s12864-015-1380-0)
- Shen X, Saburi W, Gai ZQ, Komoda K, Yu J, Ojima-Kato T, Kido Y, Matsui H, Mori H, Yao M (2014) Crystallization and preliminary x-ray crystallographic analysis of α -glucosidase HaG from *Halomonas* sp. strain H11. Acta Crystallogr F Struct Biol Commun 70:464–466. doi:[10.1107/S2053230X14001940S2053230X14001940](https://doi.org/10.1107/S2053230X14001940S2053230X14001940)
- Shirai T, Hung VS, Morinaka K, Kobayashi T, Ito S (2008) Crystal structure of GH13 α -glucosidase GSJ from one of the deepest sea bacteria. Proteins 73:126–133. doi:[10.1002/prot.22044](https://doi.org/10.1002/prot.22044)
- Sim L, Quezada-Calvillo R, Sterchi EE, Nichols BL, Rose DR (2008) Human intestinal maltase–glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. J Mol Biol 375:782–792. doi:[10.1016/j.jmb.2007.10.069](https://doi.org/10.1016/j.jmb.2007.10.069)
- Tagami T, Yamashita K, Okuyama M, Mori H, Yao M, Kimura A (2013) Molecular basis for the recognition of long-chain substrates by plant α -glucosidases. J Biol Chem 288:19296–19303. doi:[10.1074/jbc.M113.465211](https://doi.org/10.1074/jbc.M113.465211)
- Tagami T, Yamashita K, Okuyama M, Mori H, Yao M, Kimura A (2015) Structural advantage of sugar beet α -glucosidase to stabilize the Michaelis complex with long-chain substrate. J Biol Chem 290:1796–1803. doi:[10.1074/jbc.M114.606939](https://doi.org/10.1074/jbc.M114.606939)
- Wang YH, Jiang Y, Duan ZY, Shao WL, Li HZ (2009) Expression and characterization of an α -glucosidase from *Thermoanaerobacter ethanolicus* JW200 with potential for industrial application. Biologia 64:1053–1057. doi:[10.2478/s11756-009-0197-1](https://doi.org/10.2478/s11756-009-0197-1)
- Wu KY, Huang SH, Ding S, Zhang YK, Chen GG, Liang ZQ (2010) Expression, purification and characterization of recombinant α -glucosidase in *Pichia pastoris*. Folia Microbiol 66:582–587. doi:[10.1007/s12223-010-0093-7](https://doi.org/10.1007/s12223-010-0093-7)

4.3) DISCUSIÓN

Los isomaltooligosacáridos son los oligosacáridos con capacidades bifidogénicas más utilizados en los países asiáticos, con óptimas propiedades físicas de estabilidad que les convierte en mejores candidatos para su aplicación a nivel industrial como prebióticos. Actualmente, el único impedimento para considerar los IMOS como prebióticos es su digestión parcial por parte de las enzimas humanas, que degradan glucósidos con enlaces $\alpha(1-4)$. Sin embargo, este escollo se puede salvar produciendo IMOS más resistentes a la digestión intestinal humana, como aquellos que contienen enlaces $\alpha(1-2)$ y $\alpha(1-6)$, y con un mayor grado de polimerización⁴¹. Por ello, la búsqueda de nuevas enzimas con capacidad de producir IMOS es de gran interés, siendo las α -glucosidasas las mejores dianas.

La α -glucosidasa de *Xanthophyllomyces dendrorhous* (Xd- α Glu) es capaz de producir tri y tetra-IMOS con enlaces $\alpha(1-6)$. Por otra parte, también es capaz de producir kojibiosa, di-glucósido con enlace $\alpha(1-2)$. Por tanto, Xd- α Glu es una gran candidata para la síntesis de IMOS resistentes a la hidrólisis intestinal.

Como se comentó en la introducción, las α -glucosidasas de la familia GH31 que tienen alta especificidad por sustratos largos han desarrollado distintos mecanismos para dicho propósito. Algunas enzimas presentan dominios extra, como el dominio PA14 en la α -xilosidasa de *C. japonicus*⁷⁶ (figura 4.1.C), mientras que otras, han adaptado la arquitectura molecular básica GH31. Así, el largo Inserto 1 de la Ct-MGAM (figura 4.1.A) genera los subsitios +2 y +3, mientras que el N-loop de la α -glucosidasa de *B. vulgaris* (figura 4.1.B) gobierna su preferencia por sustratos largos, junto con la presencia de una secuencia concreta en el Inserto 2 que crea el resto de subsitios positivos, hasta reconocer un total de 8 unidades del malto-polisacárido⁷⁴. En base al modelo bioinformático obtenido y al análisis de sus determinantes moleculares, parece que Xd- α Glu también ha adecuado la arquitectura molecular básica GH31 a su especificidad, y presenta dos características estructurales novedosas (el Inserto N y el extenso Inserto 2), que proponemos como responsables de la preferencia por sustratos largos de esta enzima (figura 4.1.D). Por tanto, esta nueva estructura representaría un mecanismo molecular y estructural de especificidad por sustratos largos diferente, y ampliaría el conocimiento sobre los elementos implicados en la especificidad de sustrato de la familia GH31, ya de por sí muy diversa.

Sin embargo, el papel preciso que desempeñan estos bucles en la unión sólo se podrá determinar con adicionales experimentos estructurales y bioquímicos. Es por ello que actualmente se está trabajando en la expresión y purificación de Xd- α Glu para comenzar los ensayos de cristalización. Por una parte, la resolución estructural de Xd- α Glu permitirá establecer con más detalle las bases moleculares responsables de esta especificidad de producto inusual: tri y tetra-IMOS con enlaces $\alpha(1-6)$. Asimismo, sería la primera α -glucosidasa de levadura resuelta, por lo que los resultados obtenidos se podrán extrapolar a otras α -glucosidasas muy utilizadas, y estructuralmente relacionadas, como las enzimas de *S. occidentalis* y *S. cerevisiae*.

Finalmente, hay que destacar que se ha descrito la transglucosilación de compuestos interesantes para una aplicación biomédica, como la piridoxina o la hidroquinona. Así, la α -glucosidasa del molusco *Aplysia factiata* es capaz de mono- y di-glucosilar la piridoxina, uno de los tres componentes de la vitamina B6¹¹¹. Por otra parte, la α -glucosidasa de *S. cerevisiae* produce α -glucosil-hidroquinona y α -isomaltosil-hidroquinona¹¹², que aumenta su solubilidad en agua, pudiendo mejorar las propiedades farmacológicas. La hidroquinona es tóxica mientras que su forma glucósido, denominada arbutina, tiene propiedades antibacterianas y efectos de blanqueamiento de la piel¹¹³. Pensamos que la caracterización estructural completa de Xd- α Glu permitiría investigar su uso potencial para producir bioconjugados de interés.

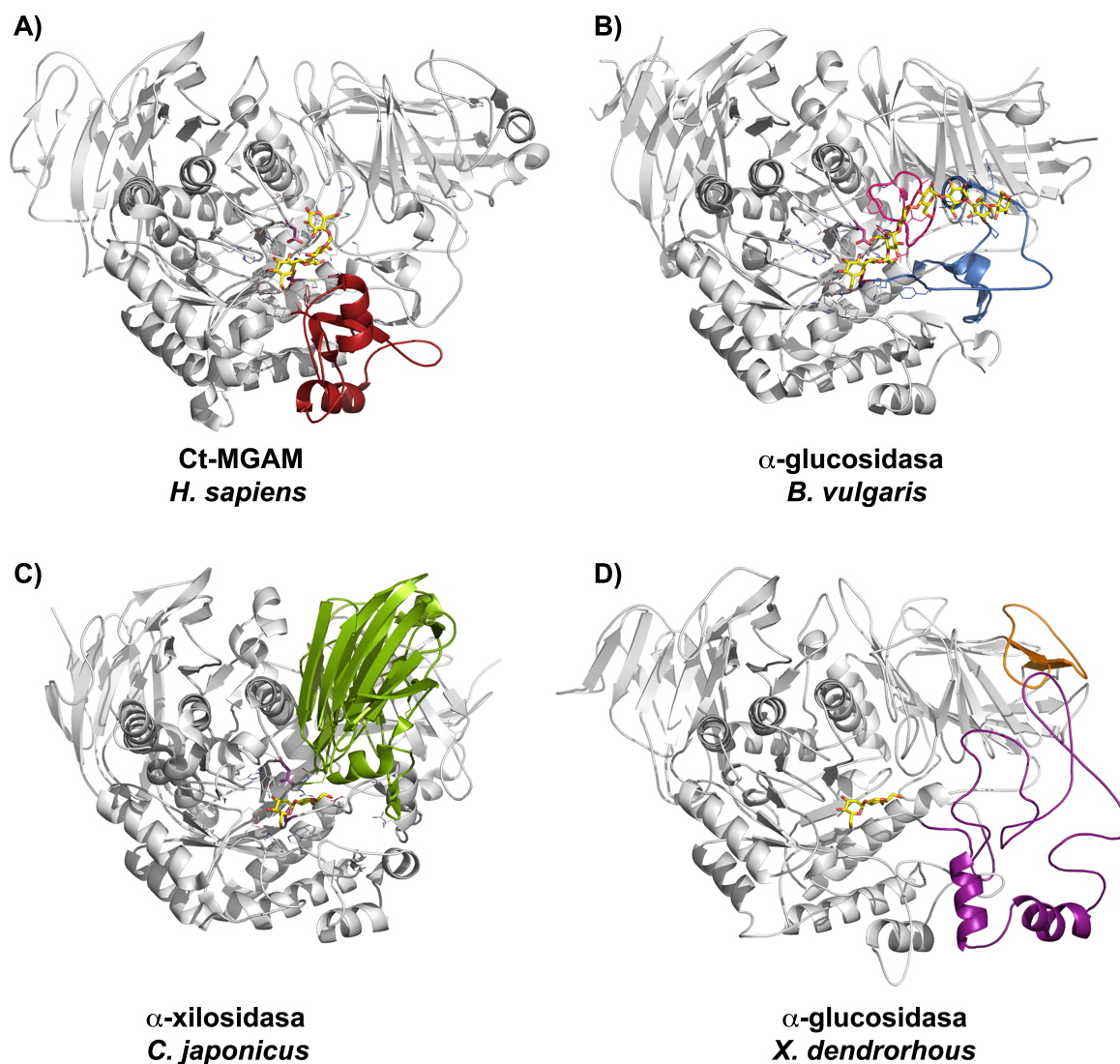


Figura 4.1. Ilustración de los determinantes estructurales de especificidad por substratos largos. (A) La maltasa-glucoamilasa C-terminal humana (Ct-MGAM) en complejo con acarbosa, resalta en rojo el Inserto 1, responsable de los subsitios +2 y +3 de la enzima. (B) La α -glucosidasa de planta, en complejo con maltoheptaosa, muestra el N-loop (rosa), que compone los subsitios +2 y +3, y el Inserto 2 (azul), que crea los subsitios +4 a +7. (C) El dominio PA14 (verde) de la α -xilosidasa bacteriana es el responsable de los sitios de reconocimiento del aglicón. (D) En el modelo estructural de Xd- α Glu se indican el Inserto N (naranja) y el Inserto2 (morado) como los determinantes estructurales responsables de la especificidad por substratos largos. En C y D se ha modelado una maltosa en el centro activo de las enzimas.

5. ENZIMAS PRODUCTORAS DE β -GLUCANOS, XOS Y GeOS: β -Glucosidasa de Rumen

5.1) INTRODUCCIÓN

Las β -glucosidasas (EC 3.2.1.21) son enzimas que liberan glucosa del extremo no reductor de los β -glucanos. Se encuentran en todos los dominios de la vida y juegan un papel fundamental en varios procesos biológicos, como el metabolismo de glicolípidos en los animales, la lignificación de la pared celular de plantas o la conversión de biomasa por microorganismos. Las β -glucosidasas se clasifican principalmente dentro de las familias GH1 y GH3, aunque también se han encontrado en otras como GH5, GH9, GH30 y GH116⁹¹.

Por su parte, las β -xilosidasas (EC 3.2.1.37) se clasifican hasta en once familias distintas (GH1, GH3, GH5, GH30, GH39, GH43, GH51, GH52, GH54, GH116 y GH120). Estas enzimas hidrolizan el enlace $\beta(1-4)$ del extremo no-reductor de los xilooligosacáridos, compuestos que provienen de la degradación de los xilanos por acción de las xilanasas. El xilano es el segundo polisacárido más abundante de la naturaleza, siendo el principal componente de la hemicelulosa¹¹⁴. En la naturaleza, estos xilooligosacáridos están frecuentemente sustituidos con acetilos, arabinofuranosidos y otras ramificaciones, por lo que los microorganismos expresan una gran variedad de enzimas específicas cuya acción combinada permite la degradación completa del complejo entramado que constituye la pared celular vegetal¹¹⁵.

Desde un punto de vista más aplicado, las β -glucosidasas y las β -xilosidasas tienen un gran interés para la conversión de biomasa ya que la degradación de la celulosa y de los componentes de la hemicelulosa son esenciales para la bioproducción de carburantes¹¹⁶. En el ámbito alimentario, las β -glucosidasas se añaden a distintos alimentos y bebidas durante el proceso de producción para mejorar sus cualidades de sabor y sus características de composición y textura. Este método puede liberar vitaminas, antioxidantes, y otros compuestos beneficiosos, de sus respectivos bioconjugados glicósidos⁹¹. Además, estas enzimas se pueden usar para mejorar las propiedades químicas y físicas de la fibra dietética²³. Asimismo, las β -glucosidasas y las β -xilosidasas pueden catalizar la reacción inversa, la transglicosilación, y sintetizar oligosacáridos y glucoconjugados de interés para la industria alimentaria y farmacéutica.

La familia GH3 es una de las más numerosas, como se comentó en la introducción de esta memoria. Hasta la fecha de escritura de este manuscrito, se han reportado numerosas estructuras tridimensionales de la familia GH3 (www.cazy.org/GH3_structure.html), de las cuales aproximadamente un tercio son β -glucosidasas, y sólo una β -xilosidasa. Estos estudios han mostrado la gran diversidad en la composición y disposición de dominios de las distintas enzimas, como se ha explicado anteriormente.

Structural and Functional Characterization of a Ruminal β -Glycosidase Defines a Novel Subfamily of Glycoside Hydrolase Family 3 with Permuted Domain Topology.

Mercedes Ramírez-Escudero, Mercedes V. del Pozo, Julia Marín-Navarro, Beatriz González, Peter N. Golyshin, Julio Polaina, Manuel Ferrer and Julia Sanz-Aparicio.

The Journal of Biological Chemistry, Vol. 291, No. 46, pp. 24200-24214,

11 de Noviembre del 2016

DOI: 10.1074/jbc.M116.747527

Códigos de acceso del PDB: 5K6L, 5K6M, 5K6N y 5K6O

Interés de la investigación:

La β -glicosidasa aislada del metagenoma de rumen bovino (GlyA₁) es una β -glucosidasa/ β -xilosidasa de amplia especificidad, que presenta además otras actividades como β -galactosidasa, β -fucosidasa, α -L-arabinosidasa y liquenasa. Esta enzima es capaz de degradar oligosacáridos de distinto grado de polimerización, siendo más activa con sustratos de tamaño medio. Estas características la convierten en una buena candidata para distintas aplicaciones biotecnológicas. Por otra parte, su mecanismo de retención y su amplia especificidad sugieren su potencial uso en la obtención de distintos prebióticos y oligosacáridos de interés biológico.

El análisis preliminar de GlyA₁ mostró una inversión en su secuencia respecto al resto de enzimas GH3 de estructura conocida. Por tanto, su estudio estructural define las características fundamentales de esta nueva subfamilia desconocida hasta el momento. Además, GlyA₁ presenta un dominio adicional en el extremo C-terminal, no descrito con anterioridad. Así, el análisis de este nuevo dominio, permitirá establecer su posible papel en la función de GlyA₁.

Los objetivos de este trabajo fueron:

- Escanear la librería de metagenoma de rumen bovino en busca de enzimas capaces de degradar β -glucanos. Aislar y clonar el gen *glyA₁*.
- Caracterizar bioquímicamente las actividades de la enzima GlyA₁, y de su versión truncada GlyA₁- Δ Ct. Determinar la especificidad por sustratos β -glucanos y β -xilanos de distinto grado de polimerización.
- Establecer las condiciones óptimas de cristalización y difracción de GlyA₁ y resolver su estructura tridimensional en complejo con β -glucosa.
- Cristalizar GlyA₁- Δ Ct y resolver su estructura en complejo con β -galactosa y β -xilosa.
- Identificar los determinantes estructurales de la especificidad y funcionalidad de GlyA₁ teniendo en cuenta los datos estructurales y bioquímicos obtenidos, así como los complejos modelados mediante *docking*.

- Analizar las diferencias estructurales entre la proteína completa GlyA₁ y la versión truncada GlyA₁-ΔCt mediante ensayos de dispersión de rayos X a bajo ángulo (SAXS).
- Caracterizar estructuralmente el dominio extra C-terminal de GlyA₁ mediante modelado bioinformático por homología de secuencia.
- Proponer un modelo estructural de la proteína completa a partir de la estructura cristalográfica, el modelo del dominio C-terminal y las curvas de dispersión de SAXS. Extrapolar la posible función del dominio C-terminal.
- Analizar las relaciones filogenéticas y estructurales de GlyA₁ con el resto de GH3 descritas, e interpretar la posible evolución estructural de los miembros de la familia.

Las conclusiones de esta investigación son:

- La proteína GlyA₁, aislada del metagenoma de rumen de vaca, es una β-glicosidasa multifuncional, que presenta principalmente actividad β-glucosidasa y β-xilosidasa por oligosacáridos de cadena media, pero también cierta actividad β-galactosidasa, β-fucosidasa, α-L-arabinofuranosidasa, α-L-arabinopiranosidasa y liquenasa. Es capaz de degradar β(1-2)-, β(1-3)-, β(1-4)-, y β(1-6)-glucobiosas.
- La estructura de GlyA₁ presenta una permutación de dominios aunque conserva la arquitectura general de las β-glucosidasas GH3. Esta alteración en la secuencia define una nueva subfamilia dentro de la familia GH3.
- Además, GlyA₁ tiene un dominio extra C-terminal, con predicción de topología β-sándwich. Los análisis de SAXS muestran que la proteína completa es una estructura compacta donde el dominio extra C-terminal se posiciona sobre el núcleo estructural de la enzima.
- Los ensayos de actividad de la proteína completa frente a la versión truncada no muestran una implicación directa del dominio C-terminal en la actividad, por lo que no parece implicado en la unión a sustratos cortos. Sin embargo, no se puede destacar que participe en la unión de sustratos poliméricos, o en la fijación de la enzima a la superficie celular.
- Los análisis filogenéticos indican que esta nueva arquitectura molecular está asociada con microorganismos que habitan en el tracto digestivo de mamíferos que se alimentan de plantas. Posiblemente, GlyA₁ procede del genoma de un microorganismo del filo *Firmicutes*, muy abundante en el tracto digestivo, y su amplia especificidad de sustrato puede estar relacionada con su presencia en el rumen bovino, donde abundan los polímeros de plantas.

La autora de esta tesis ha realizado:

- Cristalización de la proteína GlyA₁, y su versión truncada GlyA₁-ΔCt. Experimentos de *soaking* con distintos sustratos de la enzima (β-glucosa, β-xilosa o β-galactosa).
- Recogida de datos de difracción en varias instalaciones de radiación sincrotrón europeas.
- Resolución estructural de GlyA₁, GlyA₁-ΔCt, y sus complejos.
- Análisis de los resultados y estudio de los residuos determinantes de la actividad de GlyA₁.
- Toma de datos de dispersión de rayos X a bajo ángulo (SAXS) de GlyA₁ y GlyA₁-ΔCt.
- Análisis de los datos de SAXS y estudio de las diferencias estructurales entre ambas versiones de la proteína.
- Modelado del dominio extra C-terminal de GlyA₁ por homología de secuencia.
- Caracterización estructural de la proteína completa GlyA₁ mediante los resultados cristalográficos, los modelos bioinformáticos del dominio C-terminal y los datos de SAXS. Análisis de los modelos de la proteína completa y del papel del dominio extra C-terminal.

Structural and Functional Characterization of a Ruminal β -Glycosidase Defines a Novel Subfamily of Glycoside Hydrolase Family 3 with Permuted Domain Topology*

Received for publication, July 11, 2016, and in revised form, September 16, 2016. Published, JBC Papers in Press, September 27, 2016, DOI 10.1074/jbc.M116.747527

Mercedes Ramírez-Escudero[†], Mercedes V. del Pozo[§], Julia Marín-Navarro[¶], Beatriz González[‡], Peter N. Golyshin^{||**}, Julio Polaina[¶], Manuel Ferrer^{§1}, and Julia Sanz-Aparicio^{‡2}

From the [†]Department of Crystallography and Structural Biology, Institute of Physical-Chemistry "Rocasolano," Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain, the [§]Institute of Catalysis and Petrochemistry, Consejo Superior de Investigaciones Científicas, Marie Curie 2, Cantoblanco, 28049 Madrid, Spain, the [¶]Institute of Agrochemistry and Food Technology, Consejo Superior de Investigaciones Científicas, Carrer Catedratic Agustín Escardino Benlloch 7, 46980 Paterna, Valencia, Spain, the ^{||}School of Biological Sciences, Bangor University, LL57 2UW Gwynedd, United Kingdom, and the ^{**}Immanuel Kant Baltic Federal University, 236040 Kaliningrad, Russia

Edited by Gerald Hart

Metagenomics has opened up a vast pool of genes for putative, yet uncharacterized, enzymes. It widens our knowledge on the enzyme diversity world and discloses new families for which a clear classification is still needed, as is exemplified by glycoside hydrolase family-3 (GH3) proteins. Herein, we describe a GH3 enzyme (GlyA₁) from resident microbial communities in strained ruminal fluid. The enzyme is a β -glucosidase/ β -xylosidase that also shows β -galactosidase, β -fucosidase, α -arabinofuranosidase, and α -arabinopyranosidase activities. Short cello- and xylo-oligosaccharides, sophorose and gentibiose, are among the preferred substrates, with the large polysaccharide lichenan also being hydrolyzed by GlyA₁. The determination of the crystal structure of the enzyme in combination with deletion and site-directed mutagenesis allowed identification of its unusual domain composition and the active site architecture. Complexes of GlyA₁ with glucose, galactose, and xylose allowed picturing the catalytic pocket and illustrated the molecular basis of the substrate specificity. A hydrophobic platform defined by residues Trp-711 and Trp-106, located in a highly mobile loop, appears able to allocate differently β -linked bioses. GlyA₁ includes an additional C-terminal domain previously unobserved in GH3 members, but crystallization of the full-length enzyme was unsuccessful. Therefore, small angle x-ray experiments have been performed to investigate the molecular flexibility and overall putative shape. This study provided evidence

that GlyA₁ defines a new subfamily of GH3 proteins with a novel permuted domain topology. Phylogenetic analysis indicates that this topology is associated with microbes inhabiting the digestive tracts of ruminants and other animals, feeding on chemically diverse plant polymeric materials.

Family 3 of glycoside hydrolases (GH3)³ contains about 11,000 entries among which are diverse enzyme activities, including β -glucosidase, β -xylosidase, exo-chitosanase, β -N-acetylglucosaminidase, glucocerebrosidase, exo-1,4- β -glucosidase, and exo-1,3/1,4- β -glucanase, that have been characterized (1). A few reported cases are the bifunctional α -L-arabinopyranosidase/ β -galactosidase (2), N-acetyl- β -glucosaminidase/ β -glucosidase (3), β -glucosidase/cellodextrinase (4), β -xylosidase/ α -L-arabinofuranosidase (5), and β -glucosidase/ β -xylosidase (6). They are retaining enzymes that remove single glycosyl residues from the non-reducing end of their substrates. Therefore, they perform catalysis by a two-step mechanism through a covalent enzyme-glycon intermediate, which is subsequently hydrolyzed via an oxocarbenium ion-like transition state.

Despite the high number of known GH3 sequences, structural knowledge on members of the GH3 family was absent until 1999, when the three-dimensional structure of the β -D-glucan exohydrolase Exo1 from *Hordeum vulgare* (barley) was reported (7). This study showed the core structure of most GH3 enzymes consisting of an N-terminal (α/β)₈ barrel domain 1,

* This work was supported by Grants BIO2013-48779-C4-2-R, BIO2013-48779-C4-3-R, and BIO2014-54494-R from the Spanish Ministry of Economy and Competitiveness; ERA Net IB2 Project MetaCat through the Spanish Ministry of Economy and Competitiveness Grant PCIN-2014-107; United Kingdom's Biotechnology and Biological Sciences Research Council Grant BB/M029085/1; European Union's Horizon 2020 Research and Innovation Program (Blue Growth, Unlocking the potential of Seas and Oceans) Grant 634486, and the European Regional Development Fund (ERDF). The authors declare that they have no conflicts of interest with the contents of this article.

✂ Author's Choice—Final version free via Creative Commons CC-BY license. The atomic coordinates and structure factors (codes 5K6L, 5K6M, 5K6N, and 5K6O) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ To whom correspondence may be addressed. Tel.: 34-91-585-4872; E-mail: mferrer@icp.csic.es.

² To whom correspondence may be addressed. Tel.: 34-91-561-9400; E-mail: xjulia@iqfr.csic.es.

³ The abbreviations used are: GH3, glycoside hydrolase family-3; CBM, carbohydrate-binding module; DP, degree of polymerization; pNP, p-nitrophenyl; pNP α Glc, pNP- α -glucose; pNP α Mal, pNP- α -maltose; pNP β Glc, pNP- β -glucose; pNP β Cel, pNP- β -cellobiose; pNP α Araf, pNP- α -arabinofuranose; pNP β Arap, pNP- β -arabinopyranose; pNP α Xyl, pNP- α -xylose; pNP β Xyl, pNP- β -xylose; pNP β Xylb, pNP- β -xylobiose; pNP α Fuc, pNP- α -fucose; pNP α Rha, pNP- α -rhamnose; pNP α Man, pNP- α -mannose; pNP β Man, pNP- β -mannose; pNP α Gal, pNP- α -galactose; pNP β Gal, pNP- β -galactose; pNP β Lac, pNP- β -lactose; pNP α GlcNAc, pNP-N-acetyl- β -D-glucosaminide; pNP α GalNAc, pNP-N-acetyl- β -D-galactosaminide; r.m.s.d., root mean square deviation; SRF, seminal rumen fluid; SAXS, small-angle x-ray scattering; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; PDB, Protein Data Bank; FLD, fibronectin-like type III domain.

which houses the active site pocket and the nucleophile, and a C-terminal $(\alpha/\beta)_6$ -sandwich domain 2, containing the acid/base catalyst. The contribution of different domains in supplying crucial catalytic residues was a highly unusual feature of GH3 enzymes. Furthermore, in the last few years many new structural studies have shown a great variety in domain composition and arrangement of typical GH3 β -glycosidases, having up to four separate domains (8–15). Although this variety produces a shift in the sequence position of the acid/base catalyst, the known structures revealed that its structural location is well conserved among the different members. In contrast, several reported structures have revealed a more uniform pattern of the β -N-acetylglucosaminidases (NagZ) members showing that, despite a few having two domains, most Gram-negative bacteria encode single domain enzymes, and all of them have the acid/base catalyst in an unusual histidine/aspartate dyad located in a flexible loop of the $(\alpha/\beta)_8$ barrel (16). This highly mobile loop has been proved to participate in substrate distortion to a 1S_3 conformation, therefore forming a productive Michaelis complex along catalysis (17). This has not been observed in other GH3 enzymes, with the substrate being in a relaxed chair conformation, although a Michaelis complex has been recently reported for the *Listeria innocua* β -glucosidase (18). Among all GH3 β -glycosidases with available structures, insights into the substrate specificity observed in the family has been reported for the *H. vulgare* Exo1 (7, 19–21) or the β -glucosidases from *Thermotoga neapolitana* (8) and *Kluyveromyces marxianus* (9). However, the high varieties in structure and composition found among the different enzymes make it difficult to extrapolate general rules explaining function, and a clear classification of different subfamilies is still needed.

A proper classification of GH3 glycosidases may require extensive biochemical and structural characterization of new enzymes. In this context, nature provides an inexhaustible reservoir from which enzymes can be isolated (22), because they are continuously changing and evolving as a consequence of natural processes of selection. Genomics and metagenomics have made accessible such an enormous reserve of uncharacterized enzymes. Thus, we and others have recently taken advantage of sequencing and extensive screening technologies to develop enzyme discovery strategies and to identify microbial enzymes with improved and unusual activities and specificities (23–25), as well as distinct active site architectures and substrate preferences relative to other structurally characterized enzymes (26). These elegant studies demonstrated that nature contains proteins with novel and/or altered sequences and protein structures, the analysis of which represents one of the major challenges in postgenomic biology (27).

Here, activity screening of a metagenomic library created from rumen fluid led us to the isolation of a novel β -glycosidase, GlyA₁, which was assigned to the GH3 family. Detailed biochemical characterization of the new enzyme revealed its substrate specificity, whereas its sequence and crystal structure analysis revealed a novel permuted domain topology, defining a new subgroup within the GH3 family. The enzyme contains an additional C-terminal domain, previously unidentified, with its molecular flexibility being explored by small angle x-ray scattering (SAXS) analysis. The structural and biochemical analysis

of the GlyA₁ hydrolase presented in this study shed new light on comparative catalysis and evolutionary model studies as well as phylogenetic relationships.

Results

Library Screening—A subset of 14,000 clones from resident microbial communities of strained ruminal fluid (SRF) collected from rumen-fistulated, non-lactating Holstein cows (28) was screened for its ability to hydrolyze *p*-nitrophenyl- β -D-glucoside (*p*NP β Glc) and *p*-nitrophenyl- β -D-cellobioside (*p*NP β Cel). We identified a positive clone (designated SRF4) that is highly active against both substrates. The fosmid with insert SRF4 (38,710 bp; G + C 41.89%) was fully sequenced. A gene herein designated as *glyA*₁ encoding a potential GH3 β -glycosidase (GlyA₁) was identified out of the 38 distinct genes on the hit fosmid. The deduced molecular mass and estimated pI value were 101,849 Da and 4.86, respectively. This 921-amino acid-long putative protein exhibited a maximum amino acid sequence identity of 59% to a similar protein in public databases (with a top hit EDO57841.1 from *Clostridium* sp.). A search of oligonucleotide patterns against the GOHTAM database (29) and TBLASTX analysis revealed compositional similarities between the DNA fragment (38,710 bp) containing the gene for GlyA₁ with genomic sequences of *Eubacterium*, *Butyrivibrio*, and *Coproccoccus* spp. BLASTN revealed similarities of short DNA fragments to *Prevotella* and *Paenibacillus* spp. BLASTX (search by translated DNA sequences) showed similarity to glycosidases of unknown *Clostridia* (phylum Firmicutes). BLASTP search with identified protein sequences showed good matches for many of them against corresponding proteins in *Eubacterium* and *Prevotella* and members of Lachnospiraceae, *Clostridium*, *Ruminococcus*, and *Bacteroides*. Most likely, GlyA₁ has thus its origin in the phylum Firmicutes, and the presence of a phage gene may, however, indicate a horizontal gene transfer of the carbohydrate metabolism genes from Firmicutes to Bacteroidetes. Those microbes are known to be abundant in the ruminal environment and are thought to play key roles in the breakdown of proteins and carbohydrate polymers (30, 31).

Biochemical Characterization of GlyA₁—The gene encoding putative GH3 β -glycosidase (GlyA₁) was cloned, expressed in *E. coli* BL21 (DE3), and purified. The hydrolytic activity was analyzed using 18 synthetic model *p*-nitrophenyl (*p*NP) derivatives with different sugars as well as a series of 11 additional oligosaccharides. Their specific activities (units/g protein) (Table 1) and the half-saturation (Michaelis) coefficient (K_m), the catalytic rate constant (k_{cat}), and the catalytic efficiency (k_{cat}/K_m) values (Table 2) were determined. As shown in Table 1, activity was confirmed for 18 substrates that revealed that GlyA₁ is a GH3 member with clear β -glucosidase and β -xylosidase activities, but also possessing β -galactosidase, β -fucosidase, α -arabinofuranosidase, and α -arabinopyranosidase activities at low level in this order (Table 1). The activity toward *p*NP-*N*-acetyl- β -D-glucosaminide (*p*NP β GlcNAc) and *p*NP-*N*-acetyl- β -D-galactosaminide (*p*NP β GalNAc) was below detection limits, and thus the enzyme does not have β -N-acetylglucosaminidase nor β -N-acetylgalactosaminidase activity. As shown in Table 2, in terms of catalytic efficiencies, *p*NP β Cel

Structure and Function of a Ruminal β -Glycosidase

TABLE 1

Substrate specificity of the purified β -glycosidase GlyA₁ and truncated GlyA₁- Δ Ct

Substrate	Specific activity	
	GlyA ₁	GlyA ₁ - Δ Ct
	units/g	
<i>p</i> NP β Glc	2226.8 \pm 120.2	197.0 \pm 10.6
<i>p</i> NP β Xyl	2876.2 \pm 111.1	624.3 \pm 34.1
<i>p</i> NP β Xylb	301.6 \pm 8.3	14.8 \pm 0.70
<i>p</i> NP β Cel	290.7 \pm 9.9	17.04 \pm 0.50
<i>p</i> NP β Gal	11.8 \pm 2.2	0.64 \pm 0.01
<i>p</i> NP β Fuc	1.18 \pm 0.01	0.62 \pm 0.01
<i>p</i> NP α Araf	1.16 \pm 0.01	0.64 \pm 0.01
<i>p</i> NP α Arap	0.66 \pm 0.01	0.33 \pm 0.01
Cellobiose	551.5 \pm 2.6	43.6 \pm 2.2
Cellotriose	532.6 \pm 0.5	44.4 \pm 4.9
Cellotetraose	569.3 \pm 0.4	48.7 \pm 5.1
Cellopentaose	641.5 \pm 0.5	53.3 \pm 2.1
Xylobiose	634.2 \pm 4.3	115.0 \pm 4.4
Xylotriase	668.3 \pm 7.2	136.1 \pm 6.7
Xyloetraose	674.5 \pm 11.5	174.7 \pm 4.5
Xylopentaose	747.9 \pm 3.5	196.4 \pm 8.6
Gentibiose	535.2 \pm 0.7	45.9 \pm 1.2
Sophorose	602.8 \pm 0.7	54.3 \pm 4.5
Lichenan	68.6 \pm 4.8	9.8 \pm 0.7

TABLE 2

Kinetic parameters of the purified β -glycosidase GlyA₁

Substrate	K_m	k_{cat}	k_{cat}/K_m
	mM	s ⁻¹	s ⁻¹ M ⁻¹
<i>p</i> NP β Glc	10.7 \pm 2.0	1.63 \pm 0.38	152.3
<i>p</i> NP β Xyl	8.8 \pm 0.5	0.95 \pm 0.44	107.8
<i>p</i> NP β Cel	1.4 \pm 0.2	0.51 \pm 0.15	314.2
<i>p</i> NP β Xylb	2.5 \pm 0.3	0.73 \pm 0.12	292.0
<i>p</i> NP β Gal	7.6 \pm 0.1	0.13 \pm 0.04	17.1
<i>p</i> NP β Fuc	4.8 \pm 0.3	0.05 \pm 0.01	10.4
<i>p</i> NP α Araf	7.8 \pm 1.7	0.03 \pm 0.01	3.85
<i>p</i> NP α Arap	10.7 \pm 3.4	0.01 \pm 0.01	0.93
Cellobiose	2.4 \pm 0.3	0.07 \pm 0.01	28.2
Xylobiose	4.7 \pm 0.2	0.05 \pm 0.01	10.6

was the preferred substrate, mainly due to the higher affinity for this substrate as compared with other *p*NP sugars. The purified recombinant hydrolase was also assayed for their activities toward different polymeric substrates. Using specific activity determination, GlyA₁ hydrolyzed all short cello- and xylo-oligosaccharides tested (degree of polymerization (DP) from 2 to 5), with longer substrates being slightly preferred (Table 1). The catalytic efficiencies (k_{cat}/K_m) while using the non-activated substrates cellobiose and xylobiose were lower than those found for *p*NP β Cel and *p*NP β -xylobiose (*p*NP β Xylb), respectively, mainly due to a significant decrease of k_{cat} values for the natural disaccharides (Table 2). A comparison of kinetic parameters using the natural substrates xylobiose and cellobiose and the synthetic *p*NP β Xylb and *p*NP β Cel substrates confirmed the \sim 2-fold higher affinity for oligosaccharides containing β -linked glucosyl *versus* xylosyl substrates. In contrast, the affinities for the monosaccharides *p*NP β Glc and *p*NP β Xyl were essentially similar, suggesting that affinity constraints are higher as the size of the oligosaccharides increases. However, due to the differences in k_{cat} values, no major differences in catalytic performance were observed when comparing β -Xyl- and β -Glc-containing sugars. The catalytic performance (k_{cat}/K_m) found for other substrates is from low to very low mainly due to lower catalytic rates. The enzyme also exhibited activity against lichenan, suggesting that is able to hydrolyze substrates with mixed β -1,3/1,4 linkages. No activity was detected using

avicel or filter paper, as well as toward substrates without β -1,4 linkages such as β -1,3 glucan or mixed β -1,3/1,6 linkages such as laminarin. Accordingly, the enzyme showed a clear preference for short cello-oligosaccharide substrates, which may likely be produced in natural settings from the cellulose components of plant cell walls due to the action of glucanases in the ruminal fluid. Other substrates such as gentibiose (containing D-glucoses joined by a β -1,6-linkage) and sophorose (or 2-O- β -D-glucopyranosyl- α -D-glucose) were also hydrolyzed to a similar extent as cellobiose and xylobiose. The optimum activity for GlyA₁ was observed within a mesophilic range (45–65 °C) and within a neutral or slightly acid pH (6.0–7.0), being most active at 55 °C and a pH close to 6.5 (Fig. 1).

Biochemical Characterization of GlyA₁- Δ Ct—A mutant containing a missing C-terminal region, herein referred to as GlyA₁- Δ Ct, was created in the vector pQE80L. After purification, activity was determined for the 18 sugars being hydrolyzed by the wild-type enzyme, so the effect of the C-terminal region was tested. As shown in Table 1, the specific activity of the mutant was from 2- to 18.4-fold lower than that of the wild type, suggesting the importance of this region in the overall activity of the enzyme. The negative effect of the elimination of the C-terminal domain (compared with the full-length protein) was most notable for the hydrolysis of sugars containing β -glucose (from 11.3- to 17.1-fold activity reduction) as compared with those containing β -xylose (from 4.6- to 5.5-fold lower activity).

Crystal Structure Determination—Preliminary crystals from the wild-type GlyA₁ were obtained after more than 3 months with PEG3350 as the precipitant, and they were cryoprotected into 25% D-glucose to obtain the complex with this sugar. The structure was solved by molecular replacement using the domains from *T. neapolitana* β -glucosidase as independent search models. Refinement and analysis of electron density maps allowed modeling of the chain containing residues 3–798 but did not show any density to build the C-terminal segment 800–921, suggesting a putative cleavage of this region in the slow crystallization step. The low numbers of crystals impeded analysis of the intact protein by mass spectrometry, but SDS-PAGE analysis of protein solution samples revealed the presence of two bands after incubation at room temperature or treatment with proteases. Therefore, the sample was incubated with subtilisin previously to the crystallization step, which accelerated formation of many good quality crystals, under similar conditions and with the same space group. These crystals were cryoprotected into 20% glycerol, and this molecule was found bound at the active site. Furthermore, crystals from a truncated construct containing residues 2–799 (GlyA₁- Δ Ct) also grew in a week with ammonium sulfate as the precipitant and, despite having different shape, yielded the same cell and space group, which is consistent with the hypothesis that the wild-type sample was cleaved. These crystals were used to obtain the complexes with D-xylose and D-galactose. Many attempts done to crystallize the complete enzyme were unsuccessful. Also, a construct with residues 800–921, containing the isolated C-terminal region (GlyA₁-Ct), failed to crystallize. Crystallographic data and refinement statistics for the four structures here presented are given in Table 3.

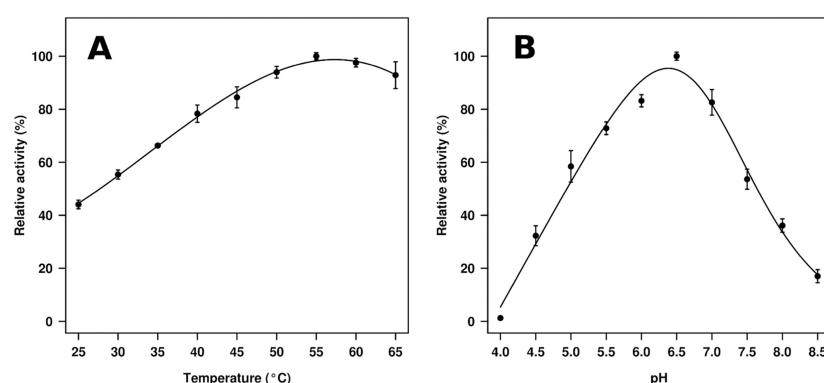


FIGURE 1. **Temperature (A) and pH (B) profiles of the purified β -glucosidase GlyA₁.** The data represent the relative percentages of specific activity (units/g) compared with the maximum activity using pNP β Glc as substrate (100% in A, 2841 units/g; 100% in B, 3056 units/g). The specific activities were calculated using 0.23 μ M protein and 10 mg/ml pNP β Glc as the assay substrate. A, reactions were performed in 50 mM sodium acetate buffer, pH 5.6, at different temperatures. B, reactions were performed at different pH (50 mM BR buffer) and at 40 °C. Standard deviations of the results of assays conducted in triplicate are shown.

TABLE 3

Crystallographic data of GlyA₁

Values in parentheses are for the high resolution shell.

Crystal data	GlyA ₁ /glycerol	GlyA ₁ /glucose	GlyA ₁ - Δ Ct/xylose	GlyA ₁ - Δ Ct/galactose
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters				
<i>a</i> (Å)	51.22	50.63	50.60	50.92
<i>b</i> (Å)	119.72	119.18	119.32	119.25
<i>c</i> (Å)	157.49	157.42	157.20	157.48
Data collection				
Beamline	Diamond (IO3)	PetraIII/DESY (P13)	ESRF (ID23-1)	ALBA (XALOC)
Temperature (K)	100	100	100	100
Wavelength (Å)	0.9762	0.9786	0.9762	1.1271
Resolution (Å)	95.31–1.83 (1.83–1.87)	95.03–2.17 (2.17–2.24)	95.05–2.08 (2.08–2.14)	95.08–2.29 (2.29–2.37)
Data processing				
Total reflections	537,914 (21,607)	338,356 (28,624)	384,429 (29,703)	287,828 (27,811)
Unique reflections	84,644 (3858)	51,199 (4369)	58,135 (4451)	44,188 (4264)
Multiplicity	6.4 (5.6)	6.6 (6.6)	6.6 (6.7)	6.5 (6.5)
Completeness (%)	98.9 (87.2)	99.7 (99.9)	99.9 (99.9)	100.0 (100.0)
Mean <i>I</i> / σ (<i>I</i>)	8.7 (2.1)	11.0 (3.3)	10.9 (3.0)	11.4 (3.3)
<i>R</i> _{merge} ^a (%)	13.7 (56.2)	12.7 (57.9)	8.7 (54.0)	9.0 (52.8)
<i>R</i> _{pim} ^b (%)	5.9 (24.9)	5.4 (24.4)	3.6 (22.4)	3.8 (22.3)
Molecules/ASU	1	1	1	1
Refinement				
<i>R</i> _{work} / <i>R</i> _{free} ^c (%)	15.68/17.85	17.48/21.82	17.38/21.25	18.16/22.72
No. of atoms/average B (Å²)				
Protein	6150/19.55	6121/31.57	6129/41.37	6151/47.72
Carbohydrate	0/0	12/38.94	10/43.17	12/53.10
Water molecules	674/28.82	328/31.23	304/41.10	120/39.00
All atoms	6878/20.66	6466/31.59	6503/41.66	6331/47.88
Ramachandran plot (%)				
Favored	98.00	97.00	98.00	98.00
Outliers	0	0	0	0
r.m.s.d.				
Bonds (Å)	0.007	0.007	0.008	0.010
Angles (°)	1.209	1.218	1.259	1.417
PDB codes				
	5K6L	5K6M	5K6N	5K6O

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement of reflection *hkl* and $\langle I(hkl) \rangle$ is the weighted mean of all measurements.

^b $R_{\text{pim}} = \sum_{hkl} (1/(N-1)) |1/2 \sum_i I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where *N* is the redundancy for the *hkl* reflection.

^c $R_{\text{work}}/R_{\text{free}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} F_o$, where F_c is the calculated and F_o is the observed structure factor amplitude of reflection *hkl* for the working/free (5%) set.

Permuted Domain Topology of GlyA₁—The first solved structure from barley β -D-glucan glucohydrolase (7) showed the core structure common to GH3 enzymes, composed of an N-terminal (α/β)₈ barrel domain 1 linked to an (α/β)₆-sandwich domain 2 (Fig. 2A); both of them provided residues that make up the active site. The later reported structures from *T. neapolitana* (8), *Trichoderma reesei* (12), *Aspergillus* (13, 14),

and *L. innocua* (18) β -glucosidases, and a β -glucosidase isolated from soil compost (32), showed the presence of an additional fibronectin type III (FnIII) domain (also designated fibronectin-like domain or FLD) located at the C terminus. This three-domain arrangement is shared by other reported β -glucosidases from *K. marxianus* (9) and *Streptomyces venezuelae* (11) that also contain an additional PA14 domain inserted

Structure and Function of a Ruminal β -Glucosidase

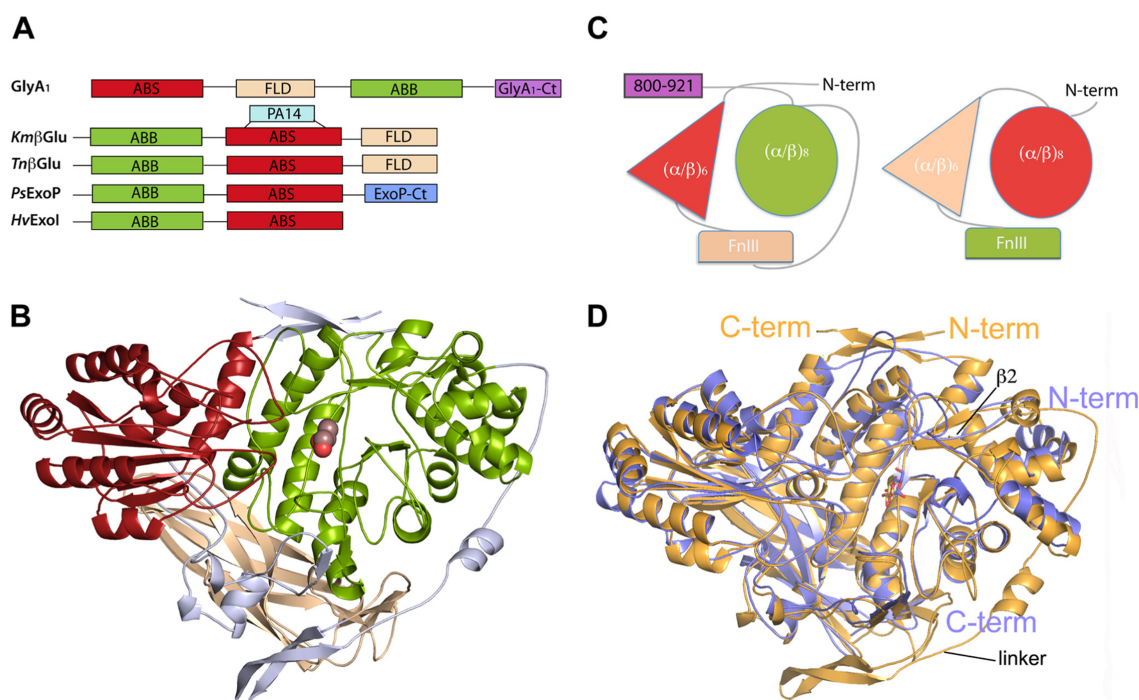


FIGURE 2. Permuted domain composition of GlyA₁. *A*, comparison of GlyA₁ structure with representative members of multidomain GH3 enzymes. β -glucosidases from *K. marxianus*, Km β Glu (9) and *T. neapolitana*, Tn β Glu (8), the exo-1,3/1,4- β -glucanase from *Pseudoalteromonas* sp., PsExoP (10) and the barley β -D-glucan exohydrolase, HvExoI (7) are shown. Domains are named as ABS: $(\alpha/\beta)_6$ -sandwich; FLD fibronectin-like; ABB $(\alpha/\beta)_8$ barrel; PA14, protective antigen PA14 domain. *B*, folding of GlyA₁. The N-terminal $(\alpha/\beta)_6$ -sandwich domain (red) is followed by the FnIII domain (beige) and the $(\alpha/\beta)_8$ barrel domain (green). Two long segments connect the three domains (gray). A glucose found in the active site is represented by spheres. *C*, scheme of the GlyA₁ domain organization (left) as compared with that of *T. neapolitana* β -glucosidase (right) (8). *D*, superimposition of GlyA₁ (gold) onto *T. neapolitana* β -glucosidase (blue) coordinates. Both enzymes present a deviation from the canonical $(\alpha/\beta)_8$ barrel topology, with their first α -helix missing, which makes strand β 2 reversed and antiparallel with the other seven strands. The main difference between both enzymes is the long arm linking the FnIII to the $(\alpha/\beta)_8$ domain in GlyA₁, which is missing in *T. neapolitana* β -glucosidase. Also, small differences in the orientation of some helices are observed.

within the same loop of their $(\alpha/\beta)_6$ -sandwich, although both are arranged in a different orientation. Moreover, the structure of the *Pseudoalteromonas* sp. exo-1,3/1,4- β -glucanase has been reported to have a C-terminal domain attached to the core structure, structurally related to family 30 carbohydrate-binding modules (CBM30), although its function is unknown (10). To expand even more this diverse landscape, GlyA₁ presents a novel structural arrangement showing permuted sequence and topology, in which the $(\alpha/\beta)_6$ sandwich (previous domain 2) is located at the N terminus and the FnIII domain is sequentially inserted between this and the $(\alpha/\beta)_8$ barrel (Fig. 2A). Additionally, a 120-residue segment attached to the C terminus most surely folded into an additional domain.

Fig. 2B displays the 3D structure of the solved 3–798 region of GlyA₁, which present overall dimensions of 85 × 65 × 45 Å. The N-terminal $(\alpha/\beta)_6$ -sandwich domain (red, residues 10–219) is followed by the FnIII domain (beige, residues 278–419) and the $(\alpha/\beta)_8$ barrel domain (green, residues 468–780). Two long segments connect the three domains (Fig. 2B, gray). Linker 1 (residues 220–277) and half of linker 2 (residues 411–443) are tightly wrapped over the core structure, whereas the rest of linker 2 (444–467) forms an extended arm that clasps the $(\alpha/\beta)_8$ barrel. Finally, the regions at the beginning and the

end of the chain are making a two-stranded β -sheet that laces the core structure at the top.

Comparative analysis using the Dali (33) server revealed that the GlyA₁ $(\alpha/\beta)_6$ -sandwich domain, containing the catalytic acid/base residue Glu-143, superimposes onto the corresponding domain from the *T. neapolitana* β -glucosidase, with a root-mean-square deviation (r.m.s.d.) of 1.6 Å for 202 eq C α positions (39% sequence identity). The same comparison with the other structurally known GH3 gives deviations in the range 2–2.5 Å (20–25% sequence identity). The FnIII domain seems more structurally conserved along the GH3 family; GlyA₁ is most similar to those in the β -glucosidases from *T. neapolitana*, with r.m.s.d. = 1.5 Å (122 residues, 39% identity), and *K. marxianus*, with r.m.s.d. = 1.6 Å (123 residues, 32% identity), but the same analysis gives values in the range 1.8–1.9 Å (21–28% sequence identity) against the other GH3 enzymes containing this domain. Finally, the $(\alpha/\beta)_8$ barrel, which contains the nucleophile Asp-709, is most similar to the corresponding domain in the β -glucosidases from *T. neapolitana* (r.m.s.d. = 1.6 Å, 285 residues, 39% identity), *K. marxianus*, (r.m.s.d. = 1.5 Å, 276 residues, 34% identity), *S. venezuelae* (r.m.s.d. = 1.7 Å, 278 residues, 32% identity), and *T. reesei* (r.m.s.d. = 2.0 Å, 278 residues, 27% identity).

Equally to GlyA₁, all these domains present a deviation from the canonical (α/β)₈ barrel topology, which was first observed in the *T. neapolitana* β -glucosidase. Thus, their first α -helix of the eight β - α motifs is missing, which has the consequence of making strand β 2 reversed and antiparallel with the other seven strands. The different deviation from the canonical topology found at this domain is consistent with the higher deviations found in the structural comparison of GlyA₁ with other GH3 enzymes, in the range 2.5–3 Å (16–20% identity).

Interestingly, the GlyA₁ core is structurally rather conserved with known β -glucosidases with equivalent domain architecture (Fig. 2C). The superposition of the *T. neapolitana* β -glucosidase onto the structure of GlyA₁ reported here shows small differences in the orientation of some of the helices (Fig. 2D). The main difference is the long arm that links the FnIII to the (α/β)₈ domain in GlyA₁, which is missing in *T. neapolitana* β -glucosidase. There are also significant differences in the loops surrounding the active site both in length and orientation, which must be related to the different substrate specificity, as commented below.

Architecture of the Active Site—The active site of GlyA₁ is located at the molecular surface, at the interface between the (α/β)₈ barrel domain, which provides the nucleophile Asp-709 and the (α/β)₆-sandwich domain, contributing to the Glu-143 acid/base catalyst (Fig. 3A). The participation of Asp-709 in substrate hydrolysis was confirmed by site-directed mutagenesis (D709A) in GlyA₁ and GlyA₁- Δ Ct, as K_m and k_{cat} values could not be determined from the data obtained due to the activity value being below the detection limit. It is a pocket of 12 Å deep with a narrow entrance 4–6 Å wide. A detailed structural comparison with the *T. neapolitana* β -glucosidase (Fig. 3A) reveals the main differences in loop conformation observed around the active site that are responsible for making a deeper catalytic pocket in GlyA₁. First, loop β 7- α 7 of the (α/β)₈ barrel, following the nucleophile Asp-709 (residues 711–726), has an 11-residue insertion that extends away from the pocket and interacts with the long segment linking the FnIII domain to the barrel, which is missing in the *T. neapolitana* β -glucosidase. Here, Arg-717 makes an ion pair with Glu-447 at the small helix located in the middle of the extended linker, which helps in stabilizing this region. An important feature of this β 7- α 7 loop is the presence of Trp-711, close to the nucleophile Asp-709, that protrudes from the surface and delineates a narrow catalytic pocket. Moreover, and despite loop β 3- α 3 (residues 536–550) being shorter in GlyA₁, Arg-538 clearly bulges into the pocket contributing to constrict it even more.

With respect to the (α/β)₆-sandwich, similarly to that observed in *T. neapolitana* β -glucosidase, this domain is shaping the active site by means of two loops, residues 139–152 containing the acid/base catalyst Glu-143 and residues 100–113 enclosing Trp-106 that clearly projects into the catalytic pocket. Interestingly, the last loop is markedly flexible as it is deduced from the fact that it could only be fully traced in the ligand-free crystal, containing only glycerol in the active site, and in the galactose-soaked crystal of the truncated form. In contrast, the crystals of the full-length and truncated forms, soaked into glucose and xylose, respectively, showed poor density that precluded tracing residues 104–107. Furthermore, the

traced loops showed significant conformational changes in the different crystals at Trp-111, coupled to a change in Phe-147 from the adjacent 139–152 loop (Fig. 3A), reinforcing its intrinsic mobility. The loop equivalent to 100–113, which is highly variable within GH3 enzymes, was proposed to be involved in recognition of large substrates from the crystal structure of *T. neapolitana* β -glucosidase, which showed some disorder that precluded tracing of a segment equivalent to that not observed in some GlyA₁ crystals. Noteworthy, the non-visible region of *T. neapolitana* β -glucosidase includes Trp-420 that, consequently, may be defining additional binding subsites, similarly to Trp-106. However, the remaining sequence is not conserved, with both Phe-147 and Trp-111 being unique to GlyA₁, and therefore, the substrate recognition mode presented by the two enzymes to accommodate the substrate may be different.

Soaking with xylose and glucose showed a clear density indicating that both sugars occupy the catalytic pocket subsite –1 in a relaxed chair conformation (Fig. 3B). This subsite is well conserved among known GH3 β -glucosidases and, with the exception of the acid base catalyst, is made up entirely by residues from the (α/β)₈ barrel domain. Thus, residues from the loops emerging from the central β -strands are making a tight net of hydrogen bonds that accommodate the glycon with all its OH groups making at least two polar interactions. The glycon moiety is located by stacking to Trp-710, and the acid base catalyst Glu-143 and the nucleophile Asp-709 interact with the O1 and O2 hydroxyls, as is expected in GH enzymes. The other residues making subsite –1 are Asp-532, Arg-597, Lys-630, His-631, Arg-641, and Tyr-677. Xylose and glucose are bound in an identical position, and the glycerol molecules observed in the ligand-free crystals are mimicking the positions occupied by C2, C3, C4, and C5 from both sugars. The additional polar interaction made by the glucose O6 hydroxyl appears consistent with the higher affinity observed in GlyA₁ toward glucosides as compared with xylosides. Thus, as shown in Table 2, the affinity for cellobiose ($K_m = 2.4 \pm 0.3$ mM) was ~2-fold higher than that for xylobiose ($K_m = 4.7 \pm 0.2$ mM). Interestingly, soaking of crystals with galactose showed that this sugar displays a semi-chair conformation at subsite –1 by flattening of the C4 atom that has the axial hydroxyl substituent (Fig. 3B, inset). In this way, galactose is accommodated by essentially the same polar interactions observed in the glucose complex, thereby explaining the activity of the enzyme on β -galactosides. However, the energy cost of getting the substrate ring distortion is reflected by the lower β -galactosidase activity, as given in Tables 1 and 2. Accordingly, the low β -fucosidase and α -arabinosidase activities must reflect some degree of deviation from the glucose-binding pattern, through ring distortion and/or loss of polar interactions, but in any case the plasticity of the catalytic site provides a notable capacity of GlyA₁ to accept different sugars (from high to low and very low specificity).

As said before, and in contrast to that observed in *T. neapolitana* β -glucosidase that presents an active site opened to the solvent with only subsite –1 being defined, more subsites are apparent in GlyA₁. To delineate a putative +1 subsite, we modeled the position of the non-hydrolysable substrate analogs thiocellobiose and thiogentibiose by structural superimposition on the previously reported experimental barley complexes

Structure and Function of a Ruminant β -Glycosidase

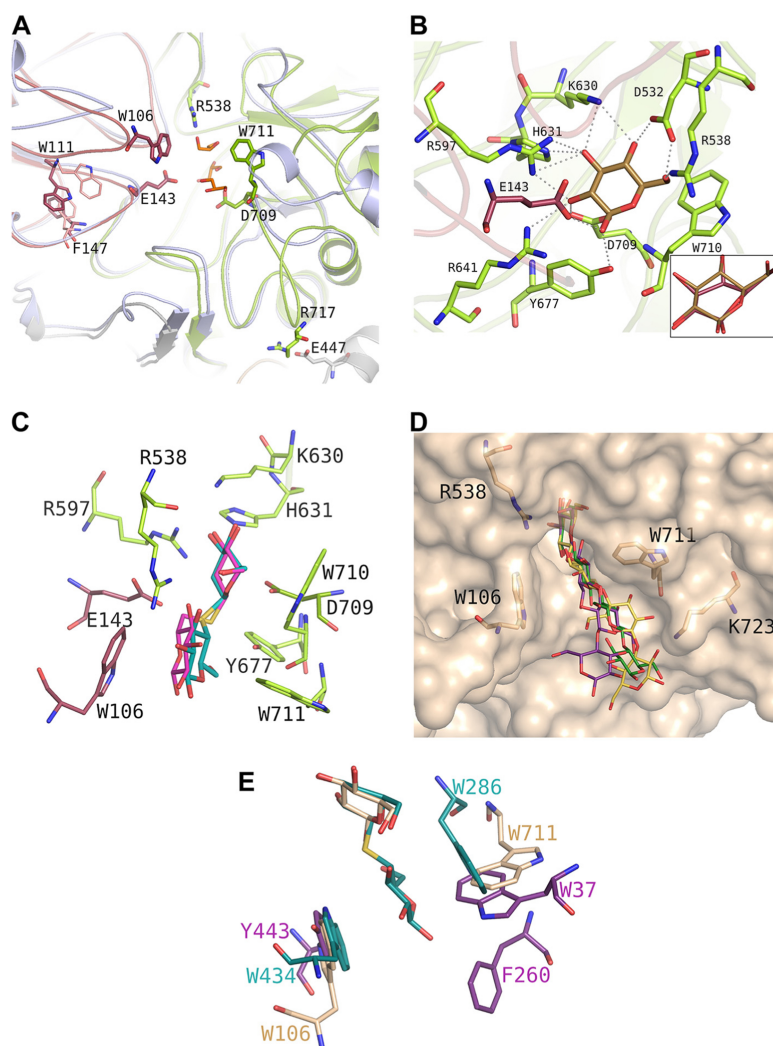


FIGURE 3. GlyA₁ active site architecture. *A*, detail of the loops surrounding its active site from the $(\alpha/\beta)_8$ barrel (green) and the $(\alpha/\beta)_8$ -sandwich (raspberry) domains, superimposed onto the *T. neapolitana* β -glucosidase (8) (pale blue). Three glycerol molecules from the cryobuffer found in the GlyA₁ crystals are shown in orange. Asp-709 and Glu-143 are the nucleophile and the acid/base catalyst, respectively. Main features of GlyA₁ are the extended loop containing Asp-709, which includes Trp-711 and the ion pair Arg-717–Glu-447 fixing it to the unique long arm and a highly flexible loop containing Trp-106. Two different conformations found among the crystals at Trp-111 and Phe-147 are highlighted. *B*, detail of the atomic interactions defining subsite –1. A glucose molecule is shown in gold. Xylose binds in the same relaxed chair conformation, and only interaction of the glucose O6 hydroxyl is missing. *Inset*, binding mode of galactose in a semi-chair conformation by flattening of the C4 atom that has the axial hydroxyl substituent and keeping the same interaction pattern. *C*, thiocellobiose (cyan) and thiogentibiose (pink) modeled at the active site by structural superimposition to the previously determined β -D-glucan glucohydrolase barley complexes (PDB entries 1IEK and 3WLP (34)), delineating putative subsite +1. *D*, molecular surface of the GlyA₁ active site, with relevant residues as sticks. Three different β -1,4/ β -1,3-linked tetraglucosides have been manually docked by superposition of their non-reduced end to the experimental glucose: a cellotetraose, as found in PDB entry 2Z1S (green); a Glc-4Glc-3Glc-4Glc (purple), and a Glc-4Glc-4Glc-3Glc (yellow), as built by the on-line carbohydrate-building program GLYCAM (45) and exported in its minimum energy state. *E*, superposition of GlyA₁-Glc structure (beige) with those reported for *T. reesei* β -glucosidase (purple) (12) and barley β -D-glucan glucohydrolase complexed with thiocellobiose (cyan) (34).

(34). As shown in Fig. 3C, Trp-106 and Trp-711 define a hydrophobic patch that may allocate the oligosaccharides at a putative subsite +1, leaving a range of possible ring orientations compatible with the observed activity of GlyA₁ against differently β -linked bioses, as given in Table 1. Also, the long chain of Arg-538, protruding at the catalytic pocket as said above, is in good position to stabilize the sugar unit by making hydrogen bonds to one or possibly two of its hydroxyl groups. The impor-

tant contribution of subsite +1 to GlyA₁ substrate binding efficiency (both glucosides and xylosides) is manifested by the lower K_m value with *pNP* β Cel compared with *pNP* β Glc and by the lower K_m value with *pNP* β Xyl compared with *pNP* β Xyl (Table 2).

Furthermore, inspection of the molecular surface of the active site cavity shown in Fig. 3D suggests the possible existence of additional subsites, which is illustrated by several

TABLE 4
SAXS data collection and derived parameters

Protein	Merged data	From Guinier			From Gnom			From Porod	
		R_g	Quality	$I(0)$	R_g	$I(0)$	D_{max}	Porod volumen	Mass
GlyA ₁ -ΔCt	mg/ml 0.34–5.29	nm 2.784	% 88	76.38	nm 2.80	76.04	nm 8.479	nm ³ 122.71	kDa 72.18
GlyA ₁	0.32–5.04	2.918	84	83.41	2.95	83.92	9.120	148.76	87.51

β -1,4/1,3-linked oligosaccharides that have been modeled at the active site as follows: a glucotetraose (green), a Glc-4Glc-3Glc-4Glc chain (purple), and a Glc-4Glc-4Glc-3Glc (yellow). These sugars have been docked by superimposition of their non-reducing units onto the observed glucose at the GlyA₁ complex. The hydrophobic patch defined by Trp-106 and Trp-711 may fit the oligosaccharides at subsites +1 and +2, and the long side chain of Lys-723 seems available to make polar interactions with the hydroxyl groups defining a possible subsite +3. The putative existence of at least three subsites in the GlyA₁ active site would be in agreement with the tendency of an increased activity against longer cello- and xylo-oligosaccharides (see Table 1). Also, the tendency of increased activity against longer cello- and xylo-oligosaccharides as given in Table 1 suggests interactions at more distal positions and therefore the possibility of additional subsites. Moreover, the shape of the active site seems compatible with the mixed β -1,4/1,3-links of the modeled tetrasaccharides, thereby explaining the observed activity on the medium size polymer lichenan.

Comparison of the GlyA₁-Glc structure with those reported for *T. reesei* β -glucosidase (12) and barley β -D-glucan glucosylase complexed with thiocellobiose (Fig. 3E) (7) displays the different hydrophobic platforms found at each active site. The barley β -D-glucan glucosylase structure showed a narrow channel with the glucose tightly arranged at subsite +1, being sandwiched between Trp-286 and Trp-434 side chains. In contrast, the GlyA₁ Trp-711 is perpendicular and oriented similarly to Trp-37 found in *T. reesei* β -glucosidase, although both residues are provided by different loops from the (α/β)₈ barrel domain. At the opposite face, GlyA₁ Trp-106 is structurally equivalent to Tyr-443 and Trp-434 from the barley and *T. reesei* enzymes, although all of them come from different loops within the (α/β)₆-sandwich domain. Interestingly, other enzymes present an aromatic residue in a position identical to Trp-106, but they are provided by the PA14 domain, Phe-508 in the case of the *K. marxianus* β -glucosidase, or by a long loop coming from the other subunit, Tyr-583 in the case of the *L. innocua* β -glucosidase dimer (data not shown) (18). This feature illustrates that these highly diverse enzymes have evolved common topology and molecular mechanisms, and yet the precise structural differences behind that regulate specificity.

SAXS Analysis of GlyA₁.—Because of the unfeasibility in crystallizing the full-length GlyA₁, we explored its overall flexibility and putative shape in solution by SAXS experiments. Thus, we compared the molecular descriptors of the complete construct with respect to the truncated construct GlyA₁-ΔCt, lacking the C-terminal domain. For this purpose, several solutions with varying concentrations were measured for each sample, and their scattering curves were merged to extrapolate idealized

data. Analysis of the scattering curves shows a good fit to the Guinier approximation, which indicates that the samples are not aggregated. Also, the calculated radii of gyration (R_g) are consistent across the range of measured concentrations. Then, the overall size descriptors can be properly determined for each construct.

First of all, the calculated molecular masses from both samples are close to the expected values (Table 4), indicating the presence of monomers, and also a 15-kDa higher mass in the complete protein, which excludes proteolysis of the analyzed sample in the short time of the experiment. Furthermore, the R_g and the maximum distance (D_{max}) for the complete protein are only slightly higher than the truncated protein, which may indicate that the extra C-terminal domain is not too extended from the core structure. In support of this hypothesis, the pairwise distance distribution function $P(r)$ calculated for both constructs shows a similar unimodal pattern consistent with a single domain protein in both cases. Furthermore, the analysis of the scattering function by the Kratki plots is consistent with the expected profile for a folded protein with a clear peak, in contrast what is observed in multidomain proteins with flexible linkers that present several peaks or smoother profiles. Consequently, we do not observe in the data calculated from the complete protein any of the signs that may be indicative of molecular flexibility, i.e. large R_g and D_{max} , absence of correlation in the $P(r)$ function, or smooth Kratki plots. Therefore, SAXS analysis appears consistent with a compact overall shape of the complete GlyA₁, in which the extra C-terminal region would not define a marked separate or flexible domain but rather it could be folded over the core three-domain structure.

To test the feasibility of this hypothesis, *ab initio* models were generated for complete GlyA₁ from SAXS data. First, two models of the last 120 residues (GlyA₁-Ct) were obtained, as explained under “Experimental Procedures,” with both showing an overall β -sandwich topology. This topology is related to carbohydrate-binding domains within families CBM6 and CBM35, to which GlyA₁-Ct presents 15–20% sequence identity, although the equivalent carbohydrate-binding motifs, typically clusters of conserved aromatic residues, are not evident in its surface. Then, three runs of CORAL were computed by considering the experimental structure of the truncated protein and each of the two models. The six models obtained are shown in Fig. 4. Analysis of these models reveals that all of them cluster around a reduced area that would locate the C-terminal region relatively distant from the catalytic pocket but quite near the mobile loop (residues 100–113). Overall, these models are consistent with the hypothesis proposed above, suggesting that GlyA₁-Ct may be somewhat packed between the two domains making the core structure and, interestingly, with a putative

Structure and Function of a Ruminal β -Glycosidase

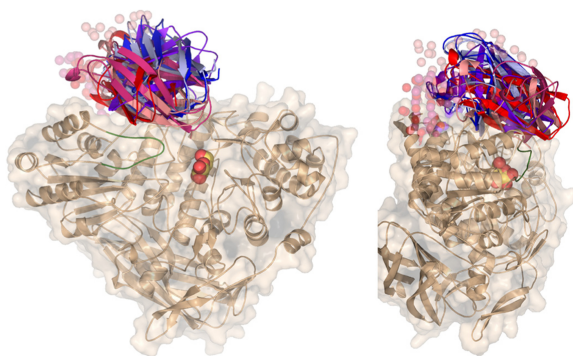


FIGURE 4. SAXS analysis of GlyA₁. Six *ab initio* models were generated for complete GlyA₁ from SAXS data, using the experimental structure of the truncated protein and two different models of the last 120 residues (GlyA₁-Ct). The two templates were obtained from Swiss-Model (red) (48) or CPHmodel (blue) (49) servers, which predict different lengths of the linker attaching this domain to the core protein, 32 or 5 residues, respectively. CORAL (47) modeling of this linker in each run is represented in spheres. The active site pocket is indicated by the galactose found at the crystal (yellow), and the mobile loop (residues 100–113), as observed in the galactose-soaked crystals, is highlighted in green.

linker somehow exposed to solvent. This feature might possibly explain the proteolysis observed in the complete protein.

GlyA₁ Phylogenetic Analysis—Our structural analysis illustrated that the permuted domain architecture of GlyA₁ keeps the location of the active site at the interface between the (α/β)₈ barrel and the (α/β)₆-sandwich domains. As mentioned above, *N*-acetylglucosaminidases are built by a single domain, with its (α/β)₈ barrel holding both the nucleophile and acid/base catalyst. Interestingly, the *Bacillus subtilis* NagZ shows the two-domain composition but still keeps the catalytic residues at the (α/β)₈ barrel (16). Therefore, this domain may be considered as the characteristic signature of GH3 enzymes. To examine the phylogenetic positioning of β -glucosidases with inverted topology (represented by GlyA₁) within the GH3 family, we have carried out a phylogenetic analysis based on the sequence of its (α/β)₈ barrel domain (ABB in this analysis). Sequences representative for each of the domain architectures found in the GH3 domain were selected (details under “Experimental Procedures”). The five topologies selected for this study are ABB, ABB-ABS, ABB-ABS-FLD, ABS-FLD-ABB, and ABB-ABS(PA14)-FLD (ABS (α/β)₆-sandwich; FLD is fibronectin-like type III domain). The resulting phylogenetic tree given in Fig. 5 shows apparent correlation between ABB sequence divergence and domain architecture. Most single domain sequences (ABB) cluster together and correspond to *N*-acetylglucosaminidases (Fig. 5, salmon area of the tree). Insertion of the ABS module is associated with three different nodes (*a*, *b*, and *c* in Fig. 5). Insertion at node *a* was not accompanied by a significant divergence in the ABB sequence because both ABB and ABB-ABS architectures appear mixed at this node. In fact, these ABB-ABS sequences also correspond to *N*-acetylglucosaminidases, and crystallographic data of *B. subtilis* NagZ show that the two modules are quite independent from a structural point of view. ABS insertion at nodes *b* and *c* would correspond to the divergence of GH3 enzymes giving rise to other activities, mainly β -glucosidase. Within node *c*, other modules (FLD and PA14)

were appended after ABS. At node *b*, fusion of C-terminal FLD seems to occur close to ABS addition because most sequences contain both modules. GlyA₁ and the other GH3 enzymes with inverted topology arose within this cluster. The phylogenetic analysis shows that the inverted topology is predominantly found in Firmicutes, although it is also present in at least another phylum (Actinobacteria) and even Archaea. Furthermore, it appears clearly associated to enzymes belonging to bacteria dwelling in the digestive tract of animals.

Discussion

In this work, a functional metagenome library analysis was used to identify a β -glycosidase from a plant polymer-degrading microorganism populating the rumen of a dairy cow. The enzyme most likely originated from the genome of a representative of Firmicutes phylum known to be abundant in the ruminal environment (30, 31).

The structural and biochemical analysis of the GlyA₁ hydrolase presented in this study sheds new light on the mechanisms of the catalysis and evolutionary patterns of the GH3 family. Our data demonstrated that GlyA₁ has a permuted domain topology. It is well documented that the formation of new domain combinations is an important mechanism in protein evolution. The major molecular mechanism that leads to multidomain proteins and novel combinations is non-homologous recombination, sometimes referred to as “domain shuffling.” This may cause recombination of domains to form different domain architectures. Proteins with the same series of domains or domain architecture are related by descent (*i.e.* evolved from one common ancestor) and tend to have the same function (35), which is rarely the case if domain order is switched. Indeed, a detailed analysis of the structures of proteins containing Rossmann fold domains demonstrated that the N- to C-terminal order of the domains is conserved because the proteins have descended from a common ancestor. For pairs of proteins in the PDB in which the order is reversed, the interface and functional relationships of the domains are altered (36). This was also proved in this study, which revealed that the altered domain architecture in GH3 mostly evolved from a distinct ecological niche, most likely from digestive tracts, including that of the ruminants. Also, the substrate specificity of the GlyA₁ protein is markedly different from that of reported GH3 members. Indeed, GlyA₁ is a uncommon multifunctional GH3 with β -glucosidase, β -xylosidase, β -galactosidase, β -fucosidase, α -arabinofuranosidase, α -arabinopyranosidase, and lichenase co-activities, with the ability to degrade β -1,2-, β -1,3-, β -1,4-, and β -1,6-glucobioses.

From an ecological point of view, the rumen compartment provides stable and favorable conditions for microbial growth and is also permanently exposed to plant biomass; for this reason, it contains specialized microorganisms that are permanently competing or collaborating for the degradation of the plant fibers. The data herein suggest that this factor, namely the high exposure to plant biomass, which is less common in other habitats, may be a strong force driving the establishment of gut microbiota with GH3 protein with permuted structures that may provide ecological advantages. Indeed, the permuted domain topology may confer the protein different functional-

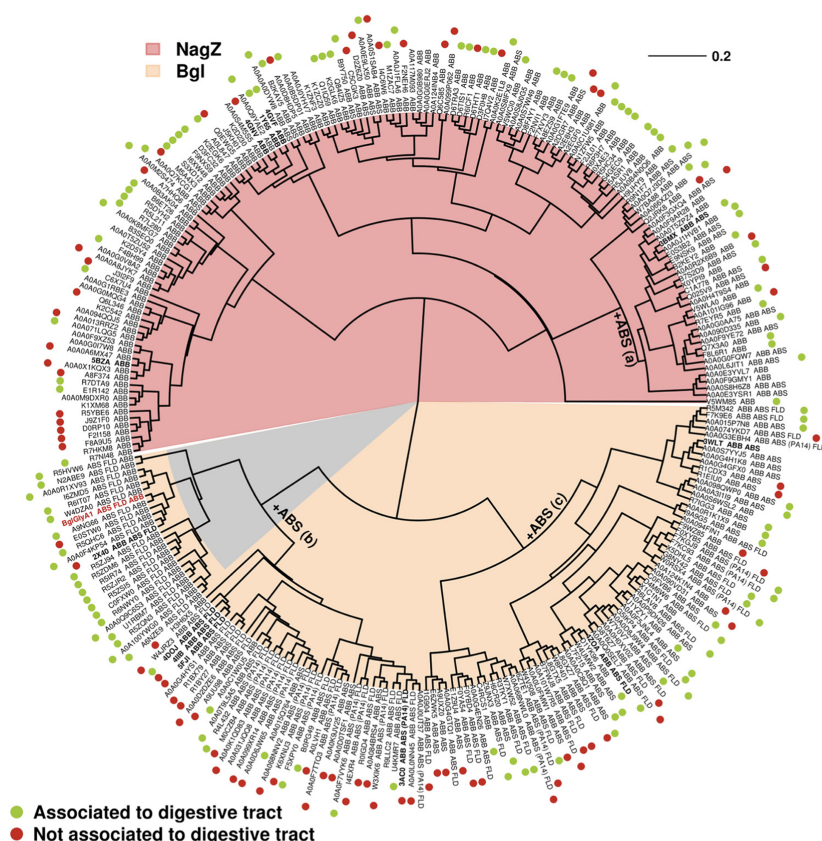


FIGURE 5. **GlyA₁ phylogenetic analysis.** The unrooted circular Neighbor-Joining tree indicating phylogenetic positions of polypeptide sequences of the GlyA₁ enzyme characterized in present work (*red boldface*) and reference similar enzymes. GenBankTM or PDB (in *boldface*) accession numbers are indicated. The domain architecture (ABB, ABS, ABB_ABS_FLD, ABB-ABS(PA14)-FLD, and ABS_FLD_ABB) to which each sequence is associated is specifically indicated. Multiple protein alignment was performed using ClustalW program, built into software version 2.1. Phylogenetic analysis was conducted with the Ape package implemented for R programming language. Sequences resembling NagZ (β -N-acetyl-glucosaminidase) are highlighted with *pink background*. Those encoding GH3 β -glucosidases are indicated in *brown*; within them, those with GlyA₁-like permuted domain topology are indicated in *gray*. ABB, (α/β)₈ barrel; ABS (α/β)₆-sandwich; FLD, fibronectin-like type III domain; PA14, protective antigen PA14 domain.

ities such as the ability to expand the pool of biomass-like substrates being hydrolyzed. Overall, our results (analysis of oligonucleotide pattern and phylogenetic tree) strongly suggest that GlyA₁ and related GH3 enzymes with inverted topology emerged in Firmicutes, where their presence is rather frequent, and are transferred by horizontal gene transfer to bacteria from other phyla and even to another kingdom (Archaea). It is well documented that these wide ranging gene transfer events take place at high frequency in the rumen (37, 38). Probably, GlyA₁ topology arose from a sequence encoding a GH3 enzyme with ABB-ABS-FLD domain architecture by gene inversion. Although the inversion surely rendered a nonfunctional gene, further mutations that would restore some sort of glycolytic activity would be strongly favored by selective pressure.

Structural analysis illustrates the permuted domain composition of GlyA₁ that is composed of an N-terminal (α/β)₆-sandwich domain, followed by the FnIII domain, and the (α/β)₈ barrel domain. Based on sequence data, a C-terminal domain was expected after the (α/β)₈ barrel domain. However, attempts to crystallize the C-terminal region of the protein

were unsuccessful, and its functional role was unclear. Biochemical characterization of the GlyA₁ and GlyA₁- Δ Ct proteins revealed that the C-terminal domain does not affect the overall substrate profile of the protein, but rather it affects the catalytic performance, which is significantly lower in the truncated GlyA₁- Δ Ct protein. This suggests that most likely the C-terminal domain may not have a direct role in substrate binding, but still it might disturb the dynamics of the proximate mobile loop (residues 100–113), which seems directly involved in catalysis.

According to available structure-prediction tools, this C-terminal region is expected to adopt a lectin-like topology, related to the CBM6/CBM35 domains. However, it does not seem an obvious carbohydrate-binding domain, and in fact, binding to xylan, cellulose, and barley glucan was not observed by affinity gel electrophoresis assays (data not shown). Nevertheless, although its involvement in binding small substrates does not seem apparent, this domain might be playing a role in positioning or locating the enzyme to distal positions of a yet unknown polymeric substrate by recognizing specific but still unidenti-

Structure and Function of a Ruminant β -Glycosidase

fied substitutions. Alternatively, it could play a role in keeping the enzyme attached to the cell surface, facilitating the intake of its products and conferring the bacteria an advantage over competing organisms. Interestingly, the analysis of the GlyA₁-Ct homologous sequences shows that these domains are attached to GH3 β -glucosidases from a ruminal environment, and this feature points to a possible function related to this ecosystem. However, its presence is not related to the permuted domain topology, as only half of the sequences included in the GlyA₁ cluster (Fig. 5) contain segments equivalent to GlyA₁-Ct.

In conclusion, the analysis of GlyA₁ here presented uncovers new features of GH3 enzymes and provides a template for a novel subfamily, including members with permuted domain topology. It also allows picturing the GlyA₁ active site architecture and the molecular basis of its substrate specificity. More work is needed to have a complete picture of the intricate molecular mechanisms that these highly diverse enzymes have evolved to tailor specificity. It will contribute to improve our knowledge about enzymatic carbohydrate degradation and open up new avenues for biocatalysis.

Experimental Procedures

Reagents and Strains—Chemicals and biochemicals were purchased from Sigma and Megazyme (Bray, Ireland) and were of pro-analysis (p.a.) quality. The oligonucleotides used for DNA amplification were synthesized by Sigma Genosys Ltd. (Pampisford, Cambs, UK). The *E. coli* Rosetta2 (Novagen, Darmstadt, Germany) for cloning and expression of wild-type protein and the genetic constructs in pQE80L vector were cultured and maintained according to the recommendations of the suppliers.

Metagenomic Library Screening and Positive-insert Sequencing—A pCC1FOS fosmid metagenomic library created from microbial communities from SRF of rumen-fistulated non-lactating Holstein cows was used. The construction and characteristics of the library were described previously (28). A subset of 14,000 clones were plated onto large (22.5 × 22.5 cm) Petri plates with Luria Bertani (LB) agar containing chloramphenicol (12.5 μ g/ml) and an arabinose-containing induction solution (Epicenter Biotechnologies) at a concentration (0.01% w/v) recommended by the supplier to induce a high fosmid copy number. After overnight incubation at 37 °C, the clones were screened for the ability to hydrolyze pNP β Glc and pNP β Cel. For screens, the plates (22.5 × 22.5 cm; each containing 2,304 clones) were covered with an agar-buffered substrate solution (40 ml of 50 mM sodium acetate buffer, pH 5.6, 0.4% w/v agar and 5 mg/ml of pNP β Glc and pNP β Cel as substrates). Positive clones were detected by the formation of a yellow color. One positive clone, herein designated as SRF4, was selected, and its DNA insert was fully sequenced with a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT) at Life Sequencing S.L. (Valencia, Spain), and the predicted genes were identified as described previously (28).

Cloning of glyA₁ and Genetic Constructs in pQE80L Plasmid—The full coding sequence of GlyA₁ (residues 2–921) and a deleted version (residues 2–799) lacking the C-terminal domain (GlyA₁- Δ Ct) were amplified by PCR with 4GF (CACGAGCTCAATATTGAAAAAGTGATACTTGATTGG) as

forward oligonucleotide and 4GR1 (AGCCGTCGACTTACTGCTGCTTTTAACTCTATTTCG) or 4GR2 (AGCCGTCGACTTACACTCTTCCTGCTATCTCAACC) as reverse oligonucleotides, respectively. The SRF4 fosmid was used as the template. The PCR conditions were as follows: 95 °C for 120 s, followed by 30 cycles of 95 °C for 30 s, 55 °C for 45 s, and 72 °C for 120 s, with a final annealing at 72 °C for 500 s. The PCR products were analyzed and agarose gel-purified using the Mini Elute gel purification kit (Qiagen, Hilden, Germany). The PCR products were digested with SacI/Sall and cloned in vector pQE80L to generate plasmids GlyA₁-pQE and GlyA₁ Δ Ct-pQE, respectively. The coding sequence of the C-terminal domain (GlyA₁-Ct, residues 800–921) was amplified with oligonucleotides CT1F (CACGAGCTCATAGAAGAGGATGCATTTCGATATAG) and 4GR1 and cloned in the SacI/Sall sites of pQE80L (plasmids Ct-pQE). GlyA₁-pQE was used as a template to introduce the mutation D709A by PCR with primers M1 (TGGTGGGCTCAGGTAAATGACC) and M2 (GGCAGTCATCACAATACCCTTAAAGCC), as described previously (39). The coding region of the resulting plasmids was fully sequenced to check for the absence of undesired mutation. The *E. coli* strain Rosetta2 (Novagen, Darmstadt, Germany) was transformed with the selected plasmids; the clones were selected on LB agar supplemented with ampicillin (100 μ g/ml) and chloramphenicol (68 μ g/ml) and stored with 20% (v/v) glycerol at –80 °C.

Site-directed Mutagenesis—Mutation D709A was introduced into the corresponding pQE80L plasmids containing genes encoding GlyA₁ and GlyA₁- Δ Ct, using the QuikChange II XL mutagenesis kit from Agilent Technologies, Inc. (Santa Clara, CA), with TGGTGGGCTCAGGTAAATGACC and GGCAGTCATCACAATACCCTTAAAGCC as forward and reverse oligonucleotides, respectively. The resulting variant plasmids were transferred into *E. coli* strain Rosetta2 (Novagen, Darmstadt, Germany) and selected on the LB agar supplemented with the same antibiotics as parental plasmids.

Gene Expression and Protein Purification—For enzyme expression and purification of wild-type and mutant GlyA₁ and GlyA₁- Δ Ct variants, as well as GlyA₁-Ct in the pQE80L vector, a single colony (*E. coli* Rosetta2) was grown overnight at 37 °C with shaking at 200 rpm in 100 ml of 2× TY medium (1% yeast extract, 1.5% tryptone, 0.5% NaCl) containing ampicillin (100 μ g/ml) and chloramphenicol (68 μ g/ml), in a 1-liter flask. Afterward, 25 ml of this culture was used to inoculate 1 liter of 2× TY medium, which was then incubated to an A_{600 nm} of ~0.6 (range from 0.55 to 0.75) at 37 °C. Protein expression was induced by 0.9 mM isopropyl β -D-galactopyranoside followed by incubation for 16 h at 16 °C. The cells were harvested by centrifugation at 5000 × g for 15 min to yield 2–3 g/liter pellet (wet weight). The cell pellet was frozen at –80 °C overnight, thawed, and resuspended in 3 ml of 20 mM phosphate buffer, pH 7.4, 500 mM NaCl/g of wet cells. Lysonase bioprocessing reagent (Novagen, Darmstadt, Germany) was then added (4 μ l/g wet cells) and incubated for 30 min on ice with rotated mixing. The cell suspension was then sonicated for a total of 1.2 min and centrifuged at 15,000 × g for 15 min at 4 °C; the supernatant was retained. The His₆-tagged enzyme was purified at 4 °C after binding to a nickel-nitrilotriacetic acid His-Bind resin

(Novagen, Darmstadt, Germany). The columns were pre-washed with 20 mM phosphate buffer, pH 7.4, 500 mM NaCl, and 50 mM imidazole, and the enzyme was eluted with the same buffer but containing 500 mM imidazole. The monitoring of the enzyme elution was performed by SDS-PAGE and/or activity measurements, using standard assays (see below). After elution, protein solution was extensively dialyzed with 20 mM Tris, pH 7.5, 50 mM NaCl by ultrafiltration through low adsorption hydrophilic 10,000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon), after which the protein was maintained at a concentration of 10 mg/ml; the protein stock solution was stored at -20°C until used in assays. The purity was assessed as $>95\%$ using SDS-PAGE, which was performed with 12% (v/v) polyacrylamide gels, using a Bio-Rad Mini Protean system. Prior to crystallization assays, 2 mM dithiothreitol (DTT) was added.

Biochemical Assays—Specific activity (units/g) and kinetic parameters (K_m and k_{cat}) were first determined using *p*NP sugars (read at 405 nm) in 96-well plates, as described previously (28). *p*NP substrates tested included those containing α -glucose (*p*NP α Glc), α -maltose (*p*NP α Mal), β -glucose (*p*NP β Glc), β -cellobiose (*p*NP β Cel), α -arabinofuranose (*p*NP α Araf), β -arabinopyranose (*p*NP β Arap), α -xylose (*p*NP α Xyl), β -xylose (*p*NP β Xyl), β -xylobiose (*p*NP β Xylb), α -fucose (*p*NP α Fuc), α -rhamnose (*p*NP α Rha), α -mannose (*p*NP α Man), β -mannose (*p*NP β Man), α -galactose (*p*NP α Gal), β -galactose (*p*NP β Gal), β -lactose (*p*NP β Lac), *N*-acetyl- β -D-glucosaminide (*p*NPGLcNAc), and *N*-acetyl- β -D-galactosaminide (*p*NPGalNAc). For cello-oligosaccharides (DP from 2 to 5), gentibiose and sophorose, the level of released glucose was determined using a glucose oxidase kit (Sigma). The level of released xylose from xylo-oligosaccharides (DP from 2 to 5) was determined using the D-xylose assay kit from Megazyme (Bray, Ireland). Substrate specificity was investigated also using carboxymethylcellulose, lichenan, barley glucan, laminarin, and avicel (all from Sigma and filter paper (Whatman, UK). Specific activity for all these sugars was quantified by measuring release of reducing sugars according to Miller (50). For K_m determinations, assay reactions were conducted by adding a protein concentration of 0.23 μM to an assay mixture containing from 0 to 30 mM sugar in 50 mM sodium acetate buffer, pH 5.6, $T = 40^{\circ}\text{C}$. Total reaction volume was 200 μL . For k_{cat} determinations, under the same conditions, sugar concentration was set up to 2 times the K_m value, and the protein concentration was from 0 to 0.23 μM . For specific activity determinations (units/g), a protein concentration of 0.23 μM and 10 mg/ml of the sugar or polysaccharide were used in 50 mM sodium acetate buffer, pH 5.6, $T = 40^{\circ}\text{C}$. The pH and temperature optima were determined in the range of pH 4.0–8.5 (50 mM Britton-Robinson buffer, BR) and 20–65 $^{\circ}\text{C}$ in assays containing a protein concentration of 0.23 μM and 10 mg/ml *p*NP β Glc, which was used as standard substrate. BR buffer is a “universal” pH buffer used for the range pH 2–12. It consists of a mixture of 0.04 M H_3BO_3 , 0.04 M H_3PO_4 , and 0.04 M CH_3COOH that has been titrated to the desired pH with 0.2 M NaOH. Optimal pH was measured at 40 $^{\circ}\text{C}$, and the optimal temperature was determined in the same buffer used in the kinetic assays. In all cases, absorbance was determined immediately after reagent and enzyme were mixed using a microplate reader every 1 min for a total time of 15 min (Synergy HT

Multi-Mode Microplate Reader, BioTek). All reactions were performed in triplicate. One unit of enzyme activity was defined as the amount of enzyme required to transform 1 μmol of substrate in 1 min under the assay conditions, with extinction coefficients as in Ref. 21. All values were corrected for non-enzymatic hydrolysis (background rate). The protein concentration was determined spectrophotometrically (at 280 nm) using a BioTek EON microplate reader (Synergy HT Multi-Mode Microplate Reader, BioTek) according to extinction coefficient of the protein (108,485 $\text{M}^{-1}\text{cm}^{-1}$) corresponding to its amino acid sequence.

Note that the detection limit, using a microplate reader with a filter for 405 nm, for the yellow chromogen is about $1 \cdot 10^{-6}$ mol/liter *p*-nitrophenol. Because the concentration of substrate in the assay ranges from 0 to 30 mM, it is expected that detection of the activity under our assay conditions is much above the detection limit.

Crystallization Data Collection and Crystal Structure Determination—Initial crystallization conditions for the complete GlyA₁ (10 mg/ml) were explored by high-throughput techniques with a NanoDrop robot (Innovadyne Technologies Inc.), using different commercial screens as follows: PACT and JCSG+ Suites from Qiagen; JBScreen Classic 1–4 from Jena Bioscience; and Index, Crystal Screen, and SaltRx packages from Hampton Research. These assays were carried out using the sitting drop vapor-diffusion method in MRC 96-well crystallization plates (Molecular Dimensions).

Elongated bars grew after 3 months in 20% polyethyleneglycol (PEG) 3350, 0.2 M ammonium sulfate, BisTris, pH 5.5. For data collection, crystals were cryoprotected in mother liquor supplemented with 25% (w/v) D-glucose before being cooled in liquid nitrogen. Diffraction data were collected at the German Electron Synchrotron (Hamburg, Germany). Diffraction images were processed with XDS (40) and scaled using Aimless from the CCP4 package (41) leading to space group $P2_12_12_1$. The structure was solved by molecular replacement using MOLREP (42) with reflections up to 2.5 Å resolution range and a Patterson radius of 54 Å. The template model was the β -glucosidase from *T. neapolitana* (PDB code 2X42), but the search was made in two steps. First, the region containing residues 2–315 was used for finding a partial solution. Then, another round of molecular replacement, with the region 321–721, was computed. Preliminary rigid-body refinement was carried out using REFMAC (43). Subsequently, several rounds of extensive model building with COOT (44) combined with automatic restraint refinement with flat bulk solvent correction and using maximum likelihood target features led to a model covering residues 3–798. However, no density was found for the loop 103–108 or for the last 123 residues of the protein. At the latter stages, β -glucose, sulfate ions, and water molecules were included in the model, which, combined with more rounds of restrained refinement, led to a final *R*-factor of 15.7 (R_{free} 17.8). The free *R*-factor was calculated using a subset of 5% randomly selected structure-factor amplitudes that were excluded from automated refinement. Many attempts to reproduce and improve these crystals were unsuccessful, until *in situ* proteolysis of the sample with subtilisin was tried. Resulting crystals grew after 15 days in the same conditions, but at pH 7.0, they were cryoprotected in 20% (v/v) glycerol and showed the same

Structure and Function of a Ruminal β -Glycosidase

space group and cell content. Then, the truncated GlyA₁- Δ Ct construct (residues 1–798) was tested. Initial crystallization assays were accomplished as described above, and several hits were obtained. Best crystals were grown in 2.0 M ammonium sulfate, 0.1 M BisTris, pH 5.5, and belonged to the same space group. The asymmetric unit contains a single molecule, with a Matthews's coefficient of 2.73 and a 54% solvent content within the cell.

Soaking experiments with D-xylose or D-galactose were performed with the truncated construct in mother liquor solution implemented with 5–50 mM ligand. Then, the crystals were flash-frozen into liquid nitrogen using mother liquor plus 20% (v/v) glycerol or ethylene glycol as cryoprotectants. The ligands were manually modeled into the electron density maps and were refined similarly to that described above. Although a mixture of α - and β -anomers may exist in solution, only the β -form of the monosaccharides was observed at the active site of the different complexes. For the docked glucotetraose coordinates, not present in the Protein Data Bank, a model was built by the on-line carbohydrate-building program GLYCAM (45).

Many attempts to crystallize the C-terminal section of the protein using the available construct were unsuccessful, and therefore, a model was built as explained below. The figures were generated with PyMOL (46). The atomic coordinates have been deposited in the RCSB Protein Data Bank under the accession codes 5K6I, 5K6M, 5K6N, and 5K6O.

SAXS Measurements—GlyA₁ and GlyA₁- Δ Ct stock solutions (10 mg/ml) were dialyzed against the same buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2 mM DTT, and 5% glycerol) for 18 h. SAXS measurements were performed at ESRF on beamline BM29, equipped with a Pilatus 1M detector. Each sample concentration, prepared by dilution of these stock solutions, was measured in 10 frames, 1-s exposure time per frame, at 4 °C, at a sample-to-detector distance of 2.867 m, using an x-ray wavelength of 0.991 Å. No radiation damage was observed during the measurements. The SAXS curves for buffer solutions were subtracted from the protein solution curves before analysis.

The scattering curves from six gradual concentrations, from 0.3 to 5 mg/ml, were scaled and averaged to obtain the $I(q)$ function using the ATSAS software package (47). The radius of gyration (R_g) for each protein was calculated by Guinier plot using the program PRIMUS, and the pair distribution function $P(r)$ and the maximum particle size D_{\max} were obtained by the program GNOM. Then, POROD was used to calculate the excluded volume of the particle, as well as the molecular weight of each sample.

Several homology and threading modeling programs were tried to obtain a model of the last 123 residues of GlyA₁. All of them predicted a topology corresponding to carbohydrate-binding domains of families CBM6/CBM35, but they differed in the length of the linker attaching this domain to the core protein. Finally, models obtained from Swiss-Model (48) and CPHmodel (49) servers were used (templates from PDB entries 2W46 and 1UYX), each predicting a loop of 32 or 5 residues, respectively. Both entries share less than 20% identity with the C-terminal region of GlyA₁.

Subsequently, CORAL (47) was used for several rounds of two-domain rigid body fitting, using the GlyA₁- Δ Ct coordinates and both templates, alternately; linkers were built as dummy atoms. The fit of the CORAL models to the SAXS

experimental data were evaluated by the χ^2 value calculated from the program CRY SOL (47).

Sequence Analysis and Construction of a Neighbor-Joining Tree—The positioning of the sequence of the GlyA₁ (α/β)₈ barrel domain was analyzed in a phylogenetic tree. The predicted protein sequences were aligned against the National Center for Biotechnology Information non-redundant (NCBI nr) database using BLASTP algorithm. We downloaded all 27,499 GH3 sequences deposited in public databases. They were grouped within five different domain architectures as follows: ABB (9, 196), ABB_ABS (3,392), ABB_ABS_FLD (11,910), ABB_ABS_PA14_FLD (2,673), and ABS_FLD_ABB (328), where ABB, ABS, FLD, and PA14 refer to (α/β)₈ barrel domain, (α/β)₆-sandwich, fibronectin-like type III domain, and protective antigen PA14 domain, respectively. We discarded those sequences (848) from the ABB_ABS group longer than 700 amino acids, as they represent enzymes with unidentified domains downstream from the ABS module. Subsequently, the sequence corresponding to the ABB domain was extracted from all of the five sub-groups. An additional filter was applied to remove ABB sequences with coverage lower than 60% of the consensus domain defined by Interpro or Pfam databases (*i.e.* with less than 200 amino acids). The final number of sequences was the following: ABB (8,109), ABB_ABS (2,312), ABB_ABS_FLD (7,335), ABB_ABS_PA14_FLD (1,664), and ABS_FLD_ABB (289). For each of the five sub-groups, redundant sequences (those sharing more than 50% identity) were eliminated to select sequences that belong to different taxonomic groups. Following this procedure, the final selected sequences were as follows: ABB (132), ABB_ABS (54), ABB_ABS_FLD (45), ABB_ABS_PA14_FLD (20), and ABS_FLD_ABB (22). Multiple protein alignment was performed using ClustalW program, built into the software version 2.1. Phylogenetic analysis was conducted with the Ape package implemented for R programming language.

Author Contributions—J. S. A., M. F., and J. P. conceived and coordinated the study. M. V. P. and M. F. contributed to screening, gene cloning, and enzyme production and characterization. P. N. G. contributed to metagenomics clone resources. J. S. A., B. G. P., and M. R. E. designed the crystallographic work and the SAXS experiments and interpreted the results. M. R. E. performed all the crystallography and SAXS experiments. J. M. N. and J. P. performed the phylogenetic analysis. J. S. A. and M. F. wrote the paper, and all authors read and commented on the manuscript.

Acknowledgments—We thank the German Electron Synchrotron (Hamburg, Germany) for assistance at Petra III P13 Beamline, the Diamond Synchrotron Radiation Source (Daresbury, UK) for assistance at I03 Beamline, and funding from the European Community's Seventh Framework Programme under BioStruct-X (Grant 283570). We also thank the staff of the European Synchrotron Radiation Facility at Grenoble (ESRF, France) for providing access and technical assistance at beamlines ID23-1 and BM29, and the Spanish Synchrotron at Barcelona (Alba, Spain) for assistance at XALOC beamline. We also acknowledge Rafael Bargiela for excellent support in the preparation of Figs. 1 and 5 and Oleg N. Reva for critical contribution to the compositional similarities analysis.

References

- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238
- Lee, J. H., Hyun, Y. J., and Kim, D. H. (2011) Cloning and characterization of α -L-arabinofuranosidase and bifunctional α -L-arabinopyranosidase/ β -D-galactopyranosidase from *Bifidobacterium longum* H-1. *J. Appl. Microbiol.* **111**, 1097–1107
- Mayer, C., Vocadlo, D. J., Mah, M., Rupitz, K., Stoll, D., Warren, R. A., and Withers, S. G. (2006) Characterization of a β -N-acetylhexosaminidase and a β -N-acetylglucosaminidase/ β -glucosidase from *Cellulomonas fimi*. *FEBS J.* **273**, 2929–2941
- DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J., and Nelson, K. E. (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J. Bacteriol.* **190**, 5455–5463
- Mai, V., Wiegel, J., and Lorenz, W. W. (2000) Cloning, sequencing, and characterization of the bifunctional xylosidase-arabinosidase from the anaerobic thermophile *Thermoanaerobacter ethanolicus*. *Gene* **247**, 137–143
- Zhou, J., Bao, L., Chang, L., Liu, Z., You, C., and Lu, H. (2012) β -Xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. *Lett. Appl. Microbiol.* **54**, 79–87
- Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) Three-dimensional structure of a barley β -D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **7**, 179–190
- Pozzo, T., Pasten, J. L., Karlsson, E. N., and Logan, D. T. (2010) Structural and functional analyses of β -glucosidase 3B from *Thermotoga neapolitana*: a thermostable three-domain representative of glycoside hydrolase 3. *J. Mol. Biol.* **397**, 724–739
- Yoshida, E., Hidaka, M., Fushinobu, S., Koyanagi, T., Minami, H., Tamaki, H., Kitaoka, M., Katayama, T., and Kumagai, H. (2010) Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β -glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431**, 39–49
- Nakatani, Y., Cutfield, S. M., Cowieson, N. P., and Cutfield, J. F. (2012) Structure and activity of exo-1,3/1,4- β -glucanase from marine bacterium *Pseudalteromonas* sp. BB1 showing a novel C-terminal domain. *FEBS J.* **279**, 464–478
- Zmudka, M. W., Thoden, J. B., and Holden, H. M. (2013) The structure of DesR from *Streptomyces venezuelae*, a β -glucosidase involved in macrolide activation. *Protein Sci.* **22**, 883–892
- Karkehabadi, S., Helmich, K. E., Kaper, T., Hansson, H., Mikkelsen, N. E., Gudmundsson, M., Piens, K., Fajdala, M., Banerjee, G., Scott-Craig, J. S., Walton, J. D., Phillips, G. N., Jr, and Sandgren, M. (2014) Biochemical characterization and crystal structures of a fungal family 3 β -glucosidase, Cel3A from *Hypocrea jecorina*. *J. Biol. Chem.* **289**, 31624–31637
- Suzuki, K., Sumitani, J., Nam, Y. W., Nishimaki, T., Tani, S., Wakagi, T., Kawaguchi, T., and Fushinobu, S. (2013) Crystal structures of glycoside hydrolase family 3 β -glucosidase 1 from *Aspergillus aculeatus*. *Biochem. J.* **452**, 211–221
- Agirre, J., Ariza, A., Offen, W. A., Turkenburg, J. P., Roberts, S. M., McNicholas, S., Harris, P. V., McBrayer, B., Dohnalek, J., Cowtan, K. D., Davies, G. J., and Wilson, K. S. (2016) Three-dimensional structures of two heavily N-glycosylated *Aspergillus* sp. family GH3 β -D-glucosidases. *Acta Crystallogr. D Struct. Biol.* **72**, 254–265
- Marín-Navarro, J., Gurgu, L., Alamar, S., and Polaina, J. (2011) Structural and functional analysis of hybrid enzymes generated by domain shuffling between *Saccharomyces cerevisiae* (var. *diastaticus*) Stal glucoamylase and *Saccharomycopsis fibuligera* Bgl1 β -glucosidase. *Appl. Microbiol. Biotechnol.* **89**, 121–130
- Litzinger, S., Fischer, S., Polzer, P., Diederichs, K., Welte, W., and Mayer, C. (2010) Structural and kinetic analysis of *Bacillus subtilis* N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *J. Biol. Chem.* **285**, 35675–35684
- Bacik, J. P., Whitworth, G. E., Stubbs, K. A., Vocadlo, D. J., and Mark, B. L. (2012) Active site plasticity within the glycoside hydrolase NagZ underlies a dynamic mechanism of substrate distortion. *Chem. Biol.* **19**, 1471–1482
- Nakajima, M., Yoshida, R., Miyanaga, A., Abe, K., Takahashi, Y., Sugimoto, N., Toyozumi, H., Nakai, H., Kitaoka, M., and Taguchi, H. (2016) Functional and structural analysis of a β -glucosidase involved in β -1,2-glucan metabolism in *Listeria innocua*. *PLoS ONE* **11**, e0148870
- Hrmova, M., De Gori, R., Smith, B. J., Fairweather, J. K., Driguez, H., Varghese, J. N., and Fincher, G. B. (2002) Structural basis for broad substrate specificity in higher plant β -D-glucan glucohydrolases. *Plant Cell* **14**, 1033–1052
- Hrmova, M., De Gori, R., Smith, B. J., Vasella, A., Varghese, J. N., and Fincher, G. B. (2004) Three-dimensional structure of the barley β -D-glucan glucohydrolase in complex with a transition state mimic. *J. Biol. Chem.* **279**, 4970–4980
- Hrmova, M., Streltsov, V. A., Smith, B. J., Vasella, A., Varghese, J. N., and Fincher, G. B. (2005) Structural rationale for low-nanomolar binding of transition state mimics to a family GH3 β -D-glucan glucohydrolase from barley. *Biochemistry* **44**, 16529–16539
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., Whitman, W. B., Euzéby, J., Amann, R., and Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645
- Alcaide, M., Tornés, J., Stogios, P. J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., Guazzaroni, M. E., López-Cortés, N., Chernikova, T. N., Golyshina, O. V., Nechitaylo, T. Y., Plumeier, I., Pieper, D. H., Yakimov, M. M., Savchenko, A., Golyshin, P. N., and Ferrer, M. (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem. J.* **454**, 157–166
- Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol.* **9**, 22–34
- Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., and Oda, K. (2016) A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351**, 1196–1199
- Alcaide, M., Stogios, P. J., Lafraya, Á., Tchigvintsev, A., Flick, R., Bargiela, R., Chernikova, T. N., Reva, O. N., Hai, T., Leggewie, C. C., Katzke, N., La Cono, V., Matesanz, R., Jebbar, M., Jaeger, K. E., et al. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol.* **17**, 332–345
- Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. (2011) The enzyme function initiative. *Biochemistry* **50**, 9950–9962
- Pozo, M. V., Fernández-Arrojo, L., Gil-Martínez, J., Montesinos, A., Chernikova, T. N., Nechitaylo, T. Y., Waliszek, A., Tortajada, M., Rojas, A., Huws, S. A., Golyshina, O. V., Newbold, C. J., Polaina, J., Ferrer, M., and Golyshin, P. N. (2012) Microbial β -glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol. Biofuels* **5**, 73
- Ménigaud, S., Mallet, L., Picord, G., Churlaud, C., Borrel, A., and Deschavanne, P. (2012) GOHTAM: a website for “Genomic Origin of Horizontal Transfers, Alignment and Metagenomics”. *Bioinformatics* **28**, 1270–1271
- Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013) Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* **7**, 1069–1079
- Pitta, D. W., Pinchak, W. E., Indugu, N., Vecchiarelli, B., Sinha, R., and Fulford, J. D. (2016) Metagenomic analysis of the rumen microbiome of steers with wheat-induced frothy bloat. *Front. Microbiol.* **7**, 689
- McAndrew, R. P., Park, J. I., Heins, R. A., Reindl, W., Friedland, G. D., D’haeseleer, P., Northen, T., Sale, K. L., Simmons, B. A., and Adams, P. D. (2013) From soil to structure, a novel dimeric β -glucosidase belonging to

Structure and Function of a Ruminant β -Glycosidase

- glycoside hydrolase family 3 isolated from compost using metagenomic analysis. *J. Biol. Chem.* **288**, 14985–14992
33. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
 34. Hrmova, M., Varghese, J. N., De Gori, R., Smith, B. J., Driguez, H., and Fincher, G. B. (2001) Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant β -D-glucan glucohydrolase. *Structure* **9**, 1005–1016
 35. Hegyi, H., and Gerstein, M. (2001) Annotation transfer for genomics: measuring functional divergence in multi-domain proteins. *Genome Res.* **11**, 1632–1640
 36. Bashton, M., and Chothia, C. (2002) The geometry of domain combination in proteins. *J. Mol. Biol.* **315**, 927–939
 37. Ricard, G., McEwan, N. R., Dutilh, B. E., Jouany, J. P., Macheboeuf, D., Mitsumori, M., McIntosh, F. M., Michalowski, T., Nagamine, T., Nelson, N., Newbold, C. J., Nsabimana, E., Takenaka, A., Thomas, N. A., Ushida, K., *et al.* (2006) Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics* **7**, 22
 38. Berg Miller, M. E., Yeoman, C. J., Chia, N., Tringe, S. G., Angly, F. E., Edwards, R. A., Flint, H. J., Lamed, R., Bayer, E. A., and White, B. A. (2012) Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ. Microbiol.* **14**, 207–227
 39. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**, 6545–6551
 40. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
 41. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
 42. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25
 43. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
 44. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
 45. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655
 46. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, version 1.6, DeLano Scientific, San Carlos, CA
 47. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350
 48. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258
 49. Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T. N. (2010) CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* **38**, W576–W581
 50. Miller, G. L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428

5.3) DISCUSIÓN

En primer lugar, esta investigación resalta la importancia de la metagenómica en la obtención de nuevas funcionalidades enzimáticas a partir de entornos no muy accesibles, pero ricos en diversidad y multiplicidad de actividades y estructuras moleculares. Este método es una estrategia útil para encontrar nuevos genes con relevancia biotecnológica.

En particular, la enzima descrita en este capítulo, GlyA₁, fue aislada de una librería de metagenoma de rumen bovino, presentando dos características únicas y originales dentro de la familia GH3, a la que pertenece. Por un lado, su secuencia de aminoácidos refleja una alteración en el orden de los dominios que conforman la estructura terciaria de la enzima, además de la presencia de una región extra C-terminal, que podría constituir otro dominio anexo. Por otro lado, el análisis de su actividad revela que es una enzima multifuncional capaz de degradar distintos tipos de carbohidratos.

Si bien ya se conocía la inversión de dominios en algunas glicosil hidrolasas bacterianas de la familia 3¹¹⁷, hasta la fecha no se había resuelto la estructura de ninguna GH3 con permutación de dominios. La inversión de secuencias es un importante mecanismo evolutivo para crear nuevas estructuras. Normalmente, la permutación en la secuencia de aminoácido de una enzima conlleva cambios apreciables en su arquitectura molecular¹¹⁸; sin embargo, GlyA₁ mantiene el patrón de plegamiento común de las β -glucosidasas de la familia GH3, representadas por la enzima de *Thermotoga neapolitana*⁸³.

Así, la estructura de GlyA₁ presentada en este trabajo representa un paradigma dentro de la familia GH3 y define una nueva subfamilia. Esta permutación de dominios conlleva algunos elementos estructurales novedosos, como el segmento conector entre el dominio FnIII y el barril (β/α)₈ que envuelve al núcleo estructural, descrito por primera vez en este trabajo de investigación. Por otra parte, dado que el centro activo de GH3 se encuentra en la interfase entre dos dominios, la predicción estructural de los residuos involucrados en la actividad en esta subfamilia no era accesible mediante modelado por homología de secuencia, utilizando como patrón las enzimas reportadas anteriormente. La estructura de GlyA₁ aquí descrita es, sin embargo, un patrón adecuado para modelar otros miembros de la familia con permutación de dominios, lo que permitirá realizar estudios de ingeniería de proteínas encaminadas al diseño de nuevas actividades en las enzimas de esta subfamilia.

La segunda característica novedosa de GlyA₁ es su amplia especificidad, ya que es una β -glucosidasa/ β -xilosidasa, pero además presenta otras actividades como β -galactosidasa, β -fucosidasa, α -L-arabinofuranosidasa, α -L-arabinopiranosidasa y liquenasa.

El análisis filogenético sugiere que la permutación de dominios está asociada con enzimas procedentes de microorganismos de la microbiota intestinal. En el tracto gastrointestinal de los animales, las bacterias que allí habitan están expuestas a distintos polisacáridos de plantas, con complejas estructuras y composiciones químicas, entre los que abundan la celulosa y la hemicelulosa. La hemicelulosa está compuesta por una red de polímeros distintos, que contienen unidades de glucosas, xilosas, fucosas, arabinosas, etc. Además, las β -xilosidasas de la familia GH3 presentan preferencia por los polisacáridos, aunque también son capaces de degradar xilooligosacáridos. Sin embargo, tienen muy poca actividad frente a xilobiosa¹¹⁹. Por tanto, GlyA₁ podría degradar diversos compuestos de las paredes celulares de plantas. Si bien es necesario realizar más experimentos para determinar la especificidad concreta frente a estos complejos polisacáridos, los ensayos bioquímicos presentes en este trabajo parecen indicar que GlyA₁ podría degradar los distintos polímeros que componen la hemicelulosa (xiloglucanos, arabinoxilanos, fucomananos, galactomananos y β -glucanos), lo que sería de gran utilidad, por ejemplo, en la

eliminación de compuestos secundarios en la industria papelera o la producción de fibra dietética más soluble y con mejores prestaciones.

Por otra parte, los β -glucanos tienen un gran interés en la industria alimentaria ya que su consumo está relacionado con la disminución del colesterol⁴³. Además, un menor grado de polimerización de estos β -glucanos parece tener un efecto más beneficioso en la salud del individuo²¹, por lo que su degradación parcial es de gran interés. Sin embargo, hasta la fecha sólo se ha explorado y reconocido el papel saludable de los β -glucanos de avena⁴⁵. Por ello creemos que la capacidad de GlyA₁ para degradar y/o modificar β -glucanos de distinto origen vegetal puede ser muy interesante para la producción de nuevas fuentes de alimentos funcionales.

Además, ya que GlyA₁ es una glicosidasa con mecanismo de retención, tiene un interesante potencial para producir distintos prebióticos emergentes por transglicosilación (figura 5.1), como xilooligosacáridos (XOS) o gentiooligosacáridos (GeOS). Los XOS son oligosacáridos formados por cadenas de D-xilosa unidas por enlaces β (1-4) que se obtienen, actualmente, por hidrólisis del xilano presente en la hemicelulosa. Estudios *in vivo* han demostrado su efecto bifidogénico¹¹, y, por tanto, son compuestos cuyo uso está en auge como prebióticos. Los GeOS son oligosacáridos de glucosas unidas por enlaces β (1-6) que no se hidrolizan en el estómago ni en el intestino delgado y producen un aumento de la población de bifidobacterias y lactobacilos¹²⁰. Se sintetizan por transglucosilación, produciendo GeOS con grados de polimerización mayores a 3¹²¹.

La capacidad de GlyA₁ para degradar oligosacáridos de β -glucanos y β -xilanos le convierte en un buen candidato para producir prebióticos emergentes de un grado de polimerización medio, más resistentes a la degradación por las enzimas intestinales humanas. En consecuencia, nuevos ensayos de transglicosilación permitirán investigar las capacidades sintéticas de esta enzima, y su posible aplicación biotecnológica.

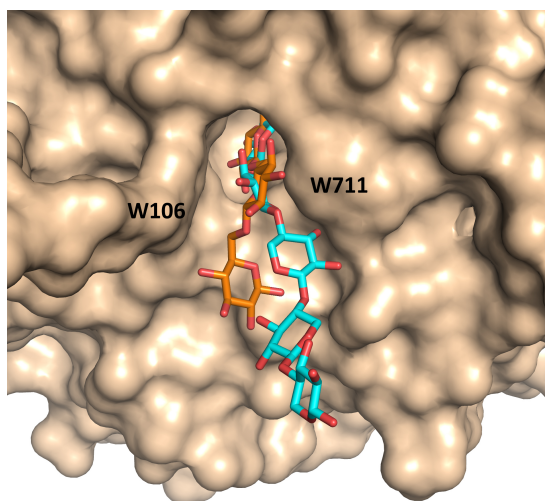


Figura 5.1 Estructura de GlyA₁ en complejo con XOS y GeOS. Sobre la superficie de GlyA₁ se han superpuesto una molécula de gentiotriosa (naranja) y una xilopentaosa (azul), por comparación con los complejos experimentales reportados en este trabajo.

6. CONSIDERACIONES FINALES

Dada la gran abundancia y variabilidad de tipos de hidratos de carbono que se incorporan en la dieta, y la escasez de genes que codifican enzimas capaces de metabolizar estos polisacáridos en el genoma humano, debemos considerar nuestras capacidades metabólicas como la suma de los constituyentes celulares microbianos y humanos⁵⁰. Además, el genoma microbiano es capaz de seguir el continuo cambio de alimentación y adaptarse⁵. Por ello, y por el creciente número de enfermedades metabólicas relacionadas con la disbiosis microbiana y la inflamación, las industrias alimentarias y farmacéuticas están incrementando su investigación en la producción de nuevos compuestos, denominados alimentos funcionales, que mejoran la salud y el bienestar humano.

Muchos de los compuestos identificados hasta el momento como prebióticos son oligosacáridos, cuyo perfil de actividad depende enormemente de su estructura química¹¹. Las enzimas que producen o procesan carbohidratos son las glicosil hidrolasas⁴⁶. Conocer sus mecanismos moleculares nos permite buscar y obtener nuevos compuestos prebióticos y realizar un diseño racional de variantes con el fin de generar biocatalizadores más eficientes. En este aspecto, la Cristalografía de rayos X aporta información a nivel atómico, vital para la mejora biotecnológica de las enzimas. Asimismo, el estudio de los complejos proteína-carbohidratos ayuda a profundizar en el conocimiento de la especificidad de sustrato inherente de cada tipo de glicosidasa, permitiendo un diseño dirigido hacia la producción de prebióticos más eficientes. Pero además, esta información estructural permite el diseño de procesos para producir novedosos compuestos bioactivos por glicosilación.

En este aspecto, aquí se presenta un estudio de cuatro enzimas GH capaces de producir carbohidratos con capacidades prebióticas: la invertasa de *S. cerevisiae* (ScInv), la β -fructofuranosidasa (XdInv) y la α -glucosidasa de *X. dendrorhous* (Xd- α Glu) y la β -glicosidasa de metagenoma de rumen bovino (GlyA₁). Sus estructuras tridimensionales han revelado los componentes estructurales implicados en el reconocimiento de los distintos sustratos, y su posible uso para la biosíntesis de distintos glicocompuestos.

Por una parte, ScInv y XdInv han demostrado capacidad de producir β (2-6)-FOS, de mayor capacidad prebiótica que los FOS actualmente comercializados. Sus estructuras tridimensionales han derivado en estudios de mutagénesis dirigida para mejorar la producción de este tipo de β (2-6)-FOS. Mientras que los mutantes de ScInv con mejoradas capacidades sintéticas que la enzima nativa han originado diversos trabajos de investigación ya publicados^{105, 122}, los ensayos de mutaciones concretas identificadas en XdInv aún se están llevando a cabo, tanto para la producción de prebióticos (neo-FOS, MFOS y otros hetero-FOS), como para la producción de fructosil-conjugados con diversos antioxidantes.

Por otro lado, el modelo computacional de la estructura de Xd- α Glu muestra dos nuevos determinantes estructurales que podrían ser los responsables de la capacidad de esta enzima de producir α (1-6)-IMOS de tamaño medio¹²³, compuestos muy interesantes ya que son más resistentes a la hidrólisis por las enzimas humanas que los comercializados siropes de IMOS, y por tanto, candidatos a ser reconocidos como prebióticos. Por ello, actualmente se está trabajando en la expresión y purificación de la enzima para comenzar con los ensayos de cristalización. Su cristalización y posterior resolución estructural permitirán dilucidar inequívocamente los residuos responsables del reconocimiento de los distintos sustratos y productos, para la síntesis dirigida de IMOS y hetero-IMOS tipo α (1-6).

Caben destacar los múltiples usos de la β -glicosidasa aislada de rumen para distintos procesos biotecnológicos. Su doble actividad β -glucosidasa/ β -xilosidasa sobre sustratos de tamaño medio,

además de sus actividades asociadas, la convierten en una buena candidata para el tratamiento de productos secundarios de ciertas industrias, como la papelera, o la conversión de biomasa. A nivel nutricional, podría ser una gran herramienta para mejorar las características de ciertos alimentos procesados, como el tratamiento de la pulpa de los zumos⁹¹, o producir fibra dietética más soluble y con mejores prestaciones para la industria alimentaria. Por último, y no menos importante, el ingente crecimiento del mercado de prebióticos hace necesario optimizar los procesos de obtención de carbohidratos prebióticos. Por ello, se está investigando en la producción de nuevos oligosacáridos a partir de los remanentes de otros procesos industriales. En este aspecto, los xilooligosacáridos y los gentiooligosacáridos son grandes candidatos ya que se producen a partir de hidrolizados de la pared celular y del almidón, las dos fuentes naturales de carbohidratos más utilizadas en industria.

Finalmente, es importante señalar que además de sus aplicaciones biotecnológicas, nuestros resultados tienen implicaciones biológicas muy relevantes. Así, hemos comprobado de manera inequívoca que el dominio no catalítico de las enzimas GH32 tiene un papel fundamental en la regulación de la estructura cuaternaria de las enzimas de levadura y, de manera directa, en su especificidad de sustrato. Es interesante resaltar que este sofisticado mecanismo de regulación no se ha encontrado en las enzimas de plantas conocidas hasta el momento, a pesar de que están implicadas en complicados procesos de crecimiento o señalización celular.

Por otra parte, hemos identificado una subfamilia dentro de la familia GH3, que presenta permutación de sus dominios y que esta asociada a los microorganismos que colonizan el tracto intestinal. Estos resultados contribuyen a tener una imagen más completa de esta amplia y compleja familia de proteínas, muchas de las cuales están implicadas en degradación de la biomasa, remodelación de la pared bacteriana o defensa frente a patógenos.

7. CONCLUSIONES

Se enumeran a continuación las principales conclusiones que se extraen del estudio de cada una de las enzimas caracterizadas en este trabajo.

Invertasa de *Saccharomyces cerevisiae* (ScInv)

1. El análisis cristalográfico de la invertasa de *Saccharomyces cerevisiae* (ScInv) muestra una estructura octamérica formada por la tetramerización de dos tipos de dímeros (abierto y cerrado), a través de la hebra β 1- β 2 del dominio β -sándwich de cada subunidad.
2. La resolución estructural de ScInv revela un monómero con topología típica GH32 y un dímero muy similares a los observados previamente en la β -fructofuranosidasa de *Schwanniomyces occidentalis* (SoFfase). Sin embargo, las diferencias en sus dominios β -sándwich tienen un papel fundamental en la formación de los distintos tipos de oligomerización, que define la especificidad de sustrato de cada enzima.
3. La configuración octamérica de ScInv impide el reconocimiento de fructooligosacáridos con más de 3-4 unidades, lo que le confiere por tanto actividad esencialmente invertasa.

β -Fructofuranosidasa de *Xanthophyllomyces dendrorhous* (XdInv)

4. La β -fructofuranosidasa de *Xanthophyllomyces dendrorhous* (XdInv) es un enzima dimérica cuya subunidad, con plegamiento típico de la familia GH32, presenta dos insertos novedosos en los extremos N- y C-terminal.
5. El dímero de XdInv muestra una topología única en la familia, con ambos centros activos en la misma cara y conectados por una hendidura. Esta disposición de los bolsillos catalíticos permitiría la degradación de fructanos mixtos ramificados.
6. El dímero de XdInv se forma a través de una N-glicosilación, situada en la interfase del dímero, y la extensión C-terminal del dominio β -sándwich, que se pliega sobre el dominio catalítico y está implicada tanto en la oligomerización como en la actividad.
7. El centro activo de XdInv presenta dos sitios de unión de sustrato, uno polar que une sustratos tipo inulina y otro hidrofóbico que une los sustratos tipo levano. Este amplio centro activo explica su amplia especificidad para transfructosilar distintos carbohidratos y otras moléculas orgánicas no-azúcares.

α -Glucosidasa de *Xanthophyllomyces dendrorhous* (Xd- α Glu)

8. La predicción estructural de la α -glucosidasa de *Xanthophyllomyces dendrorhous* (Xd- α Glu) muestra un plegamiento típico de las enzimas de eucariota de la familia GH31.
9. Este modelo bioinformático exhibe dos componentes estructurales novedosos: un inserto en el dominio N-terminal y un Inserto 2 inusualmente largo en el dominio catalítico. Proponemos que estas dos nuevas propiedades pueden ser las responsables de la preferencia por sustratos largos de Xd- α Glu, explicando la síntesis de IMOS de tamaño medio.

β -Glicosidasa aislada del metagenoma de rumen bovino (GlyA₁)

10. La estructura de la β -glicosidasa aislada del metagenoma de rumen bovino (GlyA₁) muestra una permutación de dominios aunque conserva la arquitectura general de las β -glucosidasas GH3. Esta permutación de dominios parece asociada a enzimas de microorganismos que habitan en el tracto digestivo de mamíferos que se alimentan de plantas.
11. Los cristales de GlyA₁ revelaron una versión truncada de la proteína, sin los últimos 100 residuos, que forman un dominio extra no descrito previamente en ninguna proteína. Los homólogos secuenciales más cercanos de este dominio C-terminal presentan topología de β -sándwich y son dominios de unión a carbohidratos.
12. Los análisis de dispersión de rayos X a bajo ángulo (SAXS) de la proteína en solución muestran que la enzima completa forma una estructura compacta donde el dominio extra C-terminal se encuentra previsiblemente plegado sobre el núcleo catalítico de GlyA₁.

8. BIBLIOGRAFÍA

1. Cordain, L.; Eaton, S. B.; Sebastian, A.; Mann, N.; Lindeberg, S.; Watkins, B. A.; O'Keefe, J. H.; Brand-Miller, J., Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* **2005**, *81* (2), 341-54.
2. Delzenne, N. M.; Neyrinck, A. M.; Cani, P. D., Gut microbiota and metabolic disorders: How prebiotic can work? *Br J Nutr* **2013**, *109 Suppl 2*, S81-5.
3. Rodriguez, J. M.; Murphy, K.; Stanton, C.; Ross, R. P.; Kober, O. I.; Juge, N.; Avershina, E.; Rudi, K.; Narbad, A.; Jenmalm, M. C.; Marchesi, J. R.; Collado, M. C., The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis* **2015**, *26*, 26050.
4. Cani, P. D.; Everard, A., Talking microbes: When gut bacteria interact with diet and host organs. *Mol Nutr Food Res* **2016**, *60* (1), 58-66.
5. Sonnenburg, E. D.; Sonnenburg, J. L., Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab* **2014**, *20* (5), 779-86.
6. Wiley, J. M. S., Linda; Woolverton, Christopher J.; Prescott, Lansing M., *Prescott's Microbiology*. 9th Edition ed.; McGraw Hill: New York, **2013**.
7. Arumugam, M.; Raes, J.; Pelletier, E.; Le Paslier, D.; Yamada, T.; Mende, D. R.; Fernandes, G. R.; Tap, J.; Bruls, T.; Batto, J. M.; Bertalan, M.; Borruel, N.; Casellas, F.; Fernandez, L.; Gautier, L.; Hansen, T.; Hattori, M.; Hayashi, T.; Kleerebezem, M.; Kurokawa, K.; Leclerc, M.; Levenez, F.; Manichanh, C.; Nielsen, H. B.; Nielsen, T.; Pons, N.; Poulain, J.; Qin, J.; Sicheritz-Ponten, T.; Tims, S.; Torrents, D.; Ugarte, E.; Zoetendal, E. G.; Wang, J.; Guarner, F.; Pedersen, O.; de Vos, W. M.; Brunak, S.; Dore, J.; Meta, H. I. T. C.; Antolin, M.; Artiguenave, F.; Blottiere, H. M.; Almeida, M.; Brechot, C.; Cara, C.; Chervaux, C.; Cultrone, A.; Delorme, C.; Denariáz, G.; Dervyn, R.; Foerstner, K. U.; Friss, C.; van de Guchte, M.; Guedon, E.; Haimet, F.; Huber, W.; van Hylckama-Vlieg, J.; Jamet, A.; Juste, C.; Kaci, G.; Knol, J.; Lakhdari, O.; Layec, S.; Le Roux, K.; Maguin, E.; Merieux, A.; Melo Minardi, R.; M'Rini, C.; Muller, J.; Oozeer, R.; Parkhill, J.; Renault, P.; Rescigno, M.; Sanchez, N.; Sunagawa, S.; Torrejon, A.; Turner, K.; Vandemeulebrouck, G.; Varela, E.; Winogradsky, Y.; Zeller, G.; Weissenbach, J.; Ehrlich, S. D.; Bork, P., Enterotypes of the human gut microbiome. *Nature* **2011**, *473* (7346), 174-80.
8. Petschow, B.; Dore, J.; Hibberd, P.; Dinan, T.; Reid, G.; Blaser, M.; Cani, P. D.; Degnan, F. H.; Foster, J.; Gibson, G.; Hutton, J.; Klaenhammer, T. R.; Ley, R.; Nieuwdorp, M.; Pot, B.; Relman, D.; Serazin, A.; Sanders, M. E., Probiotics, prebiotics, and the host microbiome: the science of translation. *Ann N Y Acad Sci* **2013**, *1306*, 1-17.
9. Blaut, M., Relationship of prebiotics and food to intestinal microflora. *Eur J Nutr* **2002**, *41 Suppl 1*, I11-6.
10. Koropatkin, N. M.; Cameron, E. A.; Martens, E. C., How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **2012**, *10* (5), 323-35.
11. Corzo, N.; Alonso, J. L.; Azpiroz, F.; Calvo, M. A.; Cirici, M.; Leis, R.; Lombo, F.; Mateos-Aparicio, I.; Plou, F. J.; Ruas-Madiedo, P.; Ruperez, P.; Redondo-Cuenca, A.; Sanz, M. L.; Clemente, A., [Prebiotics: concept, properties and beneficial effects]. *Nutr Hosp* **2015**, *31 Suppl 1*, 99-118.

12. Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A., Diversity of the human intestinal microbial flora. *Science* **2005**, *308* (5728), 1635-8.
13. Graf, D.; Di Cagno, R.; Fak, F.; Flint, H. J.; Nyman, M.; Saarela, M.; Watzl, B., Contribution of diet to the composition of the human gut microbiota. *Microb Ecol Health Dis* **2015**, *26*, 26164.
14. Turpin, W.; Espin-Garcia, O.; Xu, W.; Silverberg, M. S.; Kevans, D.; Smith, M. I.; Guttman, D. S.; Griffiths, A.; Panaccione, R.; Otley, A.; Xu, L.; Shestopaloff, K.; Moreno-Hagelsieb, G.; Consortium, G. E. M. P. R.; Paterson, A. D.; Croitoru, K., Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* **2016**, *48* (11), 1413-1417.
15. Cerf-Bensussan, N.; Gaboriau-Routhiau, V., The immune system and the gut microbiota: friends or foes? *Nat Rev Immunol* **2010**, *10* (10), 735-44.
16. Knights, D.; Ward, T. L.; McKinlay, C. E.; Miller, H.; Gonzalez, A.; McDonald, D.; Knight, R., Rethinking "enterotypes". *Cell host & microbe* **2014**, *16* (4), 433-7.
17. Le Chatelier, E.; Nielsen, T.; Qin, J.; Prifti, E.; Hildebrand, F.; Falony, G.; Almeida, M.; Arumugam, M.; Batto, J. M.; Kennedy, S.; Leonard, P.; Li, J.; Burgdorf, K.; Grarup, N.; Jorgensen, T.; Brandslund, I.; Nielsen, H. B.; Juncker, A. S.; Bertalan, M.; Levenez, F.; Pons, N.; Rasmussen, S.; Sunagawa, S.; Tap, J.; Tims, S.; Zoetendal, E. G.; Brunak, S.; Clement, K.; Dore, J.; Kleerebezem, M.; Kristiansen, K.; Renault, P.; Sicheritz-Ponten, T.; de Vos, W. M.; Zucker, J. D.; Raes, J.; Hansen, T.; Meta, H. I. T. c.; Bork, P.; Wang, J.; Ehrlich, S. D.; Pedersen, O., Richness of human gut microbiome correlates with metabolic markers. *Nature* **2013**, *500* (7464), 541-6.
18. Scientific concepts of functional foods in Europe. Consensus document. *Br J Nutr* **1999**, *81* Suppl 1, S1-27.
19. Shahidi, F., Nutraceuticals and functional foods: Whole versus processed foods. *Trends in Food Science & Technology* **2009**, *20* (9), 376-387.
20. Bengmark, S.; Ortiz de Urbina, J. J., Simbióticos: una nueva estrategia en el tratamiento de pacientes críticos. *Nutrición Hospitalaria* **2005**, *20*, 147-156.
21. Hamaker, B. R.; Tuncil, Y. E., A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *J Mol Biol* **2014**, *426* (23), 3838-50.
22. Jones, J. M., CODEX-aligned dietary fiber definitions help to bridge the 'fiber gap'. *Nutr J* **2014**, *13*, 34.
23. Maphosa, Y.; Jideani, V. A., Dietary fiber extraction for human nutrition—A review. *Food Reviews International* **2016**, *32* (1), 98-115.
24. Lattimer, J. M.; Haub, M. D., Effects of dietary fiber and its components on metabolic health. *Nutrients* **2010**, *2* (12), 1266-89.
25. Gibson, G. R.; Roberfroid, M. B., Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr* **1995**, *125* (6), 1401-12.
26. Tanaka, R.; Takayama, H.; Morotomi, M.; Kuroshima, T.; Ueyama, S.; Matsumoto, K.; Kuroda, A.; Mutai, M., Effects of Administration of TOS and <I>Bifidobacterium breve</I> 4006 on the Human Fecal Flora. *Bifidobacteria and Microflora* **1983**, *2* (1), 17-24.
27. Roberfroid, M., Prebiotics: the concept revisited. *J Nutr* **2007**, *137* (3 Suppl 2), 830S-7S.

28. Mussatto, S. I.; Mancilha, I. M., Non-digestible oligosaccharides: A review. *Carbohydrate Polymers* **2007**, 68 (3), 587-597.
29. Prebiotics Market Size By Ingredient (Inulin, GOS, FOS, MOS), By Application (Animal Feed, Food & Beverages [Dairy, Cereals, Baked Goods, Fermented Meat, Dry Foods], Dietary Supplements [Food, Nutrition, Infant Formulations]), Industry Analysis Report, Regional Outlook, Application Potential, Price Trends, Competitive Market Share & Forecast, 2016 – 2023. (2016) <http://www.fractovia.org/news/industry-research-report/prebiotics-market>.
30. Lewis, D. H., Nomenclature and diagrammatic representation of oligomeric fructans — a paper for discussion. *New Phytologist* **1993**, 124 (4), 583-594.
31. Sabater-Molina, M.; Larque, E.; Torrella, F.; Zamora, S., Dietary fructooligosaccharides and potential benefits on health. *J Physiol Biochem* **2009**, 65 (3), 315-28.
32. Lim, J. S.; Lee, J. H.; Kang, S. W.; Park, S. W.; Kim, S. W., Studies on production and physical properties of neo-FOS produced by co-immobilized *Penicillium citrinum* and neo-fructosyltransferase. *European Food Research and Technology* **2007**, 225 (3), 457-462.
33. Corradini, C.; Lantano, C.; Cavazza, A., Innovative analytical tools to characterize prebiotic carbohydrates of functional food interest. *Anal Bioanal Chem* **2013**, 405 (13), 4591-605.
34. Alvaro-Benito, M.; de Abreu, M.; Fernandez-Arrojo, L.; Plou, F. J.; Jimenez-Barbero, J.; Ballesteros, A.; Polaina, J.; Fernandez-Lobato, M., Characterization of a beta-fructofuranosidase from *Schwanniomyces occidentalis* with transfructosylating activity yielding the prebiotic 6-kestose. *J Biotechnol* **2007**, 132 (1), 75-81.
35. Gimeno-Perez, M.; Santos-Moriano, P.; Fernandez-Arrojo, L.; Poveda, A.; Jimenez-Barbero, J.; Ballesteros, A. O.; Fernandez-Lobato, M.; Plou, F. J., Regioselective synthesis of neo-erlose by the β -fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Process Biochemistry* **2014**, 49 (3), 423-429.
36. Semjonovs, S. P.; Maraуска, M.; Linde, R.; Grube, M.; Zikmanis, P.; Bekers, M., Development of *Bifidobacterium lactis* Bb 12 on β -(2,6)-Linked Fructan-Containing Substrate. *Engineering in Life Sciences* **2004**, 4 (5), 433-437.
37. Kilian, S.; Kritzing, S.; Rycroft, C.; Gibson, G.; du Preez, J., The effects of the novel bifidogenic trisaccharide, neokestose, on the human colonic microbiota. *World Journal of Microbiology and Biotechnology* **2002**, 18 (7), 637-644.
38. Gibson, G. R.; Ottaway, P. B.; Rastall, R., *Prebiotics: New Developments in Functional Foods*. Chandos: **2000**.
39. Marx, S. P.; Winkler, S.; Hartmeier, W., Metabolization of beta-(2,6)-linked fructose-oligosaccharides by different bifidobacteria. *FEMS Microbiol Lett* **2000**, 182 (1), 163-9.
40. Macfarlane, G. T.; Steed, H.; Macfarlane, S., Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics. *Journal of Applied Microbiology* **2008**, 104 (2), 305-344.
41. Goffin, D.; Delzenne, N.; Blecker, C.; Hanon, E.; Deroanne, C.; Paquot, M., Will isomalto-oligosaccharides, a well-established functional food in Asia, break through the European and American market? The status of knowledge on these prebiotics. *Crit Rev Food Sci Nutr* **2011**, 51 (5), 394-409.

42. Sanz, M. L.; Gibson, G. R.; Rastall, R. A., Influence of disaccharide structure on prebiotic selectivity in vitro. *J Agric Food Chem* **2005**, *53* (13), 5192-9.
43. Barsanti, L.; Passarelli, V.; Evangelista, V.; Frassanito, A. M.; Gualtieri, P., Chemistry, physico-chemistry and applications linked to biological activities of beta-glucans. *Nat Prod Rep* **2011**, *28* (3), 457-66.
44. Bach Knudsen, K. E., Microbial Degradation of Whole-Grain Complex Carbohydrates and Impact on Short-Chain Fatty Acids and Health. *Advances in Nutrition: An International Review Journal* **2015**, *6* (2), 206-213.
45. Cloetens, L.; Ulmius, M.; Johansson-Persson, A.; Akesson, B.; Onning, G., Role of dietary beta-glucans in the prevention of the metabolic syndrome. *Nutr Rev* **2012**, *70* (8), 444-58.
46. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **2014**, *42* (Database issue), D490-5.
47. Davies, G.; Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure* **1995**, *3* (9), 853-9.
48. Henrissat, B., A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **1991**, *280* (Pt 2), 309-16.
49. Davies, G. J.; Sinnott, M. L., Sorting the diverse: the sequence-based classifications of carbohydrate-active enzymes. *Biochemical Journal* **2008**.
50. El Kaoutari, A.; Armougom, F.; Gordon, J. I.; Raoult, D.; Henrissat, B., The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **2013**, *11* (7), 497-504.
51. Cantarel, B. L.; Lombard, V.; Henrissat, B., Complex carbohydrate utilization by the healthy human microbiome. *PLoS One* **2012**, *7* (6), e28742.
52. Turnbaugh, P. J.; Henrissat, B.; Gordon, J. I., Viewing the human microbiome through three-dimensional glasses: integrating structural and functional studies to better define the properties of myriad carbohydrate-active enzymes. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2010**, *66* (Pt 10), 1261-4.
53. Zechel, D. L.; Withers, S. G., Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* **2000**, *33* (1), 11-8.
54. Ardevol, A.; Rovira, C., Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J Am Chem Soc* **2015**, *137* (24), 7528-47.
55. Davies, G. J.; Wilson, K. S.; Henrissat, B., Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **1997**, *321* (Pt 2), 557-9.
56. Alberto, F.; Bignon, C.; Sulzenbacher, G.; Henrissat, B.; Czjzek, M., The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J Biol Chem* **2004**, *279* (18), 18903-10.
57. Paoli, M., Protein folds propelled by diversity. *Prog Biophys Mol Biol* **2001**, *76* (1-2), 103-30.

58. Lammens, W.; Le Roy, K.; Schroeven, L.; Van Laere, A.; Rabijns, A.; Van den Ende, W., Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. *J Exp Bot* **2009**, *60* (3), 727-40.
59. Verhaest, M.; Van den Ende, W.; Roy, K. L.; De Ranter, C. J.; Laere, A. V.; Rabijns, A., X-ray diffraction structure of a plant glycosyl hydrolase family 32 protein: fructan 1-exohydrolase IIa of *Cichorium intybus*. *Plant J* **2005**, *41* (3), 400-11.
60. Alvaro-Benito, M.; Polo, A.; Gonzalez, B.; Fernandez-Lobato, M.; Sanz-Aparicio, J., Structural and kinetic analysis of *Schwanniomyces occidentalis* invertase reveals a new oligomerization pattern and the role of its supplementary domain in substrate binding. *J Biol Chem* **2010**, *285* (18), 13930-41.
61. Verhaest, M.; Lammens, W.; Le Roy, K.; De Coninck, B.; De Ranter, C. J.; Van Laere, A.; Van den Ende, W.; Rabijns, A., X-ray diffraction structure of a cell-wall invertase from *Arabidopsis thaliana*. *Acta Crystallogr D Biol Crystallogr* **2006**, *62* (Pt 12), 1555-63.
62. Pouyez, J.; Mayard, A.; Vandamme, A. M.; Roussel, G.; Perpete, E. A.; Wouters, J.; Housen, I.; Michaux, C., First crystal structure of an endo-inulinase, INU2, from *Aspergillus ficuum*: discovery of an extra-pocket in the catalytic domain responsible for its endo-activity. *Biochimie* **2012**, *94* (11), 2423-30.
63. Reddy, A.; Maley, F., Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase. *J Biol Chem* **1996**, *271* (24), 13953-7.
64. Reddy, V. A.; Maley, F., Identification of an active-site residue in yeast invertase by affinity labeling and site-directed mutagenesis. *J Biol Chem* **1990**, *265* (19), 10817-20.
65. Nagem, R. A.; Rojas, A. L.; Golubev, A. M.; Korneeva, O. S.; Eneyskaya, E. V.; Kulminskaya, A. A.; Neustroev, K. N.; Polikarpov, I., Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. *J Mol Biol* **2004**, *344* (2), 471-80.
66. Alberto, F.; Jordi, E.; Henrissat, B.; Czjzek, M., Crystal structure of inactivated *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose. *Biochem J* **2006**, *395* (3), 457-62.
67. Lovering, A. L.; Lee, S. S.; Kim, Y. W.; Withers, S. G.; Strynadka, N. C., Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *J Biol Chem* **2005**, *280* (3), 2105-15.
68. Tagami, T.; Yamashita, K.; Okuyama, M.; Mori, H.; Yao, M.; Kimura, A., Molecular basis for the recognition of long-chain substrates by plant alpha-glucosidases. *J Biol Chem* **2013**, *288* (26), 19296-303.
69. Tan, K.; Tesar, C.; Wilton, R.; Keigher, L.; Babnigg, G.; Joachimiak, A., Novel alpha-glucosidase from human gut microbiome: substrate specificities and their switch. *FASEB J* **2010**, *24* (10), 3939-49.
70. Miyazaki, T.; Ishizaki, Y.; Ichikawa, M.; Nishikawa, A.; Tonozuka, T., Structural and biochemical characterization of novel bacterial alpha-galactosidases belonging to glycoside hydrolase family 31. *Biochem J* **2015**, *469* (1), 145-58.
71. Speciale, G.; Jin, Y.; Davies, G. J.; Williams, S. J.; Goddard-Borger, E. D., YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat Chem Biol* **2016**, *12* (4), 215-7.

72. Rozeboom, H. J.; Yu, S.; Madrid, S.; Kalk, K. H.; Zhang, R.; Dijkstra, B. W., Crystal structure of alpha-1,4-glucan lyase, a unique glycoside hydrolase family member with a novel catalytic mechanism. *J Biol Chem* **2013**, *288* (37), 26764-74.
73. Sim, L.; Quezada-Calvillo, R.; Sterchi, E. E.; Nichols, B. L.; Rose, D. R., Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. *J Mol Biol* **2008**, *375* (3), 782-92.
74. Okuyama, M.; Saburi, W.; Mori, H.; Kimura, A., alpha-Glucosidases and alpha-1,4-glucan lyases: structures, functions, and physiological actions. *Cell Mol Life Sci* **2016**, *73* (14), 2727-51.
75. Hemsworth, G. R.; Thompson, A. J.; Stepper, J.; Sobala, L. F.; Coyle, T.; Larsbrink, J.; Spadiut, O.; Goddard-Borger, E. D.; Stubbs, K. A.; Brumer, H.; Davies, G. J., Structural dissection of a complex *Bacteroides ovatus* gene locus conferring xyloglucan metabolism in the human gut. *Open Biol* **2016**, *6* (7).
76. Larsbrink, J.; Izumi, A.; Ibatullin, F. M.; Nakhai, A.; Gilbert, H. J.; Davies, G. J.; Brumer, H., Structural and enzymatic characterization of a glycoside hydrolase family 31 alpha-xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification. *Biochem J* **2011**, *436* (3), 567-80.
77. Ernst, H. A.; Lo Leggio, L.; Willemoes, M.; Leonard, G.; Blum, P.; Larsen, S., Structure of the *Sulfolobus solfataricus* alpha-glucosidase: implications for domain conservation and substrate recognition in GH31. *J Mol Biol* **2006**, *358* (4), 1106-24.
78. Chaudet, M. M.; Rose, D. R., Divergent evolution for diverse substrate recognition by family 31 glycoside hydrolases. *Biochem Cell Biol* **2016**, *94* (4), 323-30.
79. Lin, A. H.; Nichols, B. L.; Quezada-Calvillo, R.; Avery, S. E.; Sim, L.; Rose, D. R.; Naim, H. Y.; Hamaker, B. R., Unexpected high digestion rate of cooked starch by the Ct-maltase-glucoamylase small intestine mucosal alpha-glucosidase subunit. *PLoS One* **2012**, *7* (5), e35473.
80. Caputo, A. T.; Alonzi, D. S.; Marti, L.; Reca, I. B.; Kiappes, J. L.; Struwe, W. B.; Cross, A.; Basu, S.; Lowe, E. D.; Darlot, B.; Santino, A.; Roversi, P.; Zitzmann, N., Structures of mammalian ER alpha-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc Natl Acad Sci U S A* **2016**, *113* (32), E4630-8.
81. Varghese, J. N.; Hrmova, M.; Fincher, G. B., Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **1999**, *7* (2), 179-90.
82. Stubbs, K. A.; Balcewich, M.; Mark, B. L.; Vocadlo, D. J., Small molecule inhibitors of a glycoside hydrolase attenuate inducible AmpC-mediated beta-lactam resistance. *J Biol Chem* **2007**, *282* (29), 21382-91.
83. Pozzo, T.; Pasten, J. L.; Karlsson, E. N.; Logan, D. T., Structural and functional analyses of beta-glucosidase 3B from *Thermotoga neapolitana*: a thermostable three-domain representative of glycoside hydrolase 3. *J Mol Biol* **2010**, *397* (3), 724-39.
84. Yoshida, E.; Hidaka, M.; Fushinobu, S.; Koyanagi, T.; Minami, H.; Tamaki, H.; Kitaoka, M.; Katayama, T.; Kumagai, H., Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 beta-glucosidase from *Kluyveromyces marxianus*. *Biochem J* **2010**, *431* (1), 39-49.
85. Leahy, D. J.; Aukhil, I.; Erickson, H. P., 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **1996**, *84* (1), 155-64.

86. Goyal, K.; Jo Kim, B.; Kim, J. D.; Kim, Y. K.; Kitaoka, M.; Hayashi, K., Enhancement of transglycosylation activity by construction of chimeras between mesophilic and thermophilic beta-glucosidase. *Arch Biochem Biophys* **2002**, *407* (1), 125-34.
87. Zmudka, M. W.; Thoden, J. B.; Holden, H. M., The structure of DesR from *Streptomyces venezuelae*, a beta-glucosidase involved in macrolide activation. *Protein Sci* **2013**, *22* (7), 883-92.
88. Rigden, D. J.; Mello, L. V.; Galperin, M. Y., The PA14 domain, a conserved all-beta domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem Sci* **2004**, *29* (7), 335-9.
89. Nakatani, Y.; Cutfield, S. M.; Cowieson, N. P.; Cutfield, J. F., Structure and activity of exo-1,3/1,4-beta-glucanase from marine bacterium *Pseudoalteromonas* sp. BB1 showing a novel C-terminal domain. *FEBS J* **2012**, *279* (3), 464-78.
90. Litzinger, S.; Fischer, S.; Polzer, P.; Diederichs, K.; Welte, W.; Mayer, C., Structural and kinetic analysis of *Bacillus subtilis* N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *J Biol Chem* **2010**, *285* (46), 35675-84.
91. Ketudat Cairns, J. R.; Esen, A., beta-Glucosidases. *Cell Mol Life Sci* **2010**, *67* (20), 3389-405.
92. Convenio sobre la diversidad biológica. Biológica, S. d. C. s. l. D., Ed. Rio de Janeiro, 1992.
93. P., K., Rainbow code of biotechnology. *CHEMIK* **2012**, *66* (8), 811-816.
94. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012**, *485* (7397), 185-94.
95. Davids, T.; Schmidt, M.; Bottcher, D.; Bornscheuer, U. T., Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr Opin Chem Biol* **2013**, *17* (2), 215-20.
96. Vijn, I.; Smeekens, S., Fructan: more than a reserve carbohydrate? *Plant Physiol* **1999**, *120* (2), 351-60.
97. Velazquez-Hernandez, M. L.; Baizabal-Aguirre, V. M.; Bravo-Patino, A.; Cajero-Juarez, M.; Chavez-Moctezuma, M. P.; Valdez-Alarcon, J. J., Microbial fructosyltransferases and the role of fructans. *J Appl Microbiol* **2009**, *106* (6), 1763-78.
98. Ritsema, T.; Smeekens, S., Fructans: beneficial for plants and humans. *Curr Opin Plant Biol* **2003**, *6* (3), 223-30.
99. Diez-Municio, M.; de las Rivas, B.; Jimeno, M. L.; Munoz, R.; Moreno, F. J.; Herrero, M., Enzymatic synthesis and characterization of fructooligosaccharides and novel maltosylfructosides by inulosucrase from *Lactobacillus gasseri* DSM 20604. *Appl Environ Microbiol* **2013**, *79* (13), 4129-40.
100. Lammens, W.; Le Roy, K.; Yuan, S.; Vergauwen, R.; Rabijs, A.; Van Laere, A.; Strelkov, S. V.; Van den Ende, W., Crystal structure of 6-SST/6-SFT from *Pachysandra terminalis*, a plant fructan biosynthesizing enzyme in complex with its acceptor substrate 6-kestose. *Plant J* **2012**, *70* (2), 205-19.
101. Alvaro-Benito, M.; Sainz-Polo, M. A.; Gonzalez-Perez, D.; Gonzalez, B.; Plou, F. J.; Fernandez-Lobato, M.; Sanz-Aparicio, J., Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* beta-fructofuranosidase, an enzyme that produces prebiotics. *J Biol Chem* **2012**, *287* (23), 19674-86.

102. Bujacz, A.; Jedrzejczak-Krzepkowska, M.; Bielecki, S.; Redzyna, I.; Bujacz, G., Crystal structures of the apo form of beta-fructofuranosidase from *Bifidobacterium longum* and its complex with fructose. *FEBS J* **2011**, *278* (10), 1728-44.
103. Le Roy, K.; Verhaest, M.; Rabijns, A.; Clerens, S.; Van Laere, A.; Van den Ende, W., N-glycosylation affects substrate specificity of chicory fructan 1-exohydrolase: evidence for the presence of an inulin binding cleft. *New Phytol* **2007**, *176* (2), 317-24.
104. Nagae, M.; Yamaguchi, Y., Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* **2012**, *13* (7), 8398-429.
105. Marin-Navarro, J.; Talens-Perales, D.; Polaina, J., One-pot production of fructooligosaccharides by a *Saccharomyces cerevisiae* strain expressing an engineered invertase. *Appl Microbiol Biotechnol* **2015**, *99* (6), 2549-55.
106. de Abreu, M.; Alvaro-Benito, M.; Sanz-Aparicio, J.; Plou, F. J.; Fernandez-Lobato, M.; Alcalde, M., Synthesis of 6-Kestose using an Efficient β -Fructofuranosidase Engineered by Directed Evolution. *Advanced Synthesis & Catalysis* **2013**, *355* (9), 1698-1702.
107. Linde, D.; Rodriguez-Colinas, B.; Estevez, M.; Poveda, A.; Plou, F. J.; Fernandez Lobato, M., Analysis of neofructooligosaccharides production mediated by the extracellular beta-fructofuranosidase from *Xanthophyllomyces dendrorhous*. *Bioresour Technol* **2012**, *109*, 123-30.
108. Seibel, J.; Moraru, R.; Gotze, S.; Buchholz, K.; Na'amnieh, S.; Pawlowski, A.; Hecht, H. J., Synthesis of sucrose analogues and the mechanism of action of *Bacillus subtilis* fructosyltransferase (levansucrase). *Carbohydr Res* **2006**, *341* (14), 2335-49.
109. Gantt, R. W.; Peltier-Pain, P.; Thorson, J. S., Enzymatic methods for glyco(diversification/randomization) of drugs and small molecules. *Nat Prod Rep* **2011**, *28* (11), 1811-53.
110. Chiba, S. a. M., N., α -glucosidases. In *Handbook of amylases and related enzymes*, Japan, T. a. r. s. o., Ed. Pergamon Press: Oxford, **1988**; pp 104-116.
111. Tramice, A.; Giordano, A.; Andreotti, G.; Mollo, E.; Trincone, A., High-yielding enzymatic α -glucosylation of pyridoxine by marine α -glucosidase from *Aplysia fasciata*. *Mar Biotechnol (NY)* **2006**, *8* (5), 448-52.
112. Prodanović, R. M.; Milosavić, N. B.; Sladić, D.; Veličković, T. Ć.; Vujčić, Z., Synthesis of hydroquinone- α -glucoside by α -glucosidase from baker's yeast. *Biotechnology Letters* **2005**, *27* (8), 551-554.
113. Funayama, M.; Arakawa, H.; Yamamoto, R.; Nishino, T.; Shin, T.; Murao, S., Effects of α - and β -arbutin on activity of tyrosinases from mushroom and mouse melanoma. *Biosci Biotechnol Biochem* **1995**, *59* (1), 143-4.
114. Collins, T.; Gerday, C.; Feller, G., Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev* **2005**, *29* (1), 3-23.
115. Dodd, D.; Kiyonari, S.; Mackie, R. I.; Cann, I. K., Functional diversity of four glycoside hydrolase family 3 enzymes from the rumen bacterium *Prevotella bryantii* B14. *J Bacteriol* **2010**, *192* (9), 2335-45.
116. Simmons, A. B., 6.9 Hemicellulases. In *Chemical and biochemical catalysis for next generation biofuels.*, Simmons, B., Ed. Royal Society of Chemistry: **2011**.

117. Harvey, A. J.; Hrmova, M.; De Gori, R.; Varghese, J. N.; Fincher, G. B., Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins* **2000**, *41* (2), 257-69.
118. Bashton, M.; Chothia, C., The geometry of domain combination in proteins. *J Mol Biol* **2002**, *315* (4), 927-39.
119. Lagaert, S.; Pollet, A.; Courtin, C. M.; Volckaert, G., beta-xylosidases and alpha-L-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. *Biotechnol Adv* **2014**, *32* (2), 316-32.
120. Rycroft, C. E.; Jones, M. R.; Gibson, G. R.; Rastall, R. A., Fermentation properties of gentio-oligosaccharides. *Lett Appl Microbiol* **2001**, *32* (3), 156-61.
121. Kothari, D.; Goyal, A., Gentio-oligosaccharides from *Leuconostoc mesenteroides* NRRL B-1426 dextranucrase as prebiotics and as a supplement for functional foods with anti-cancer properties. *Food Funct* **2015**, *6* (2), 604-11.
122. Lafraya, A.; Sanz-Aparicio, J.; Polaina, J.; Marin-Navarro, J., Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. *Appl Environ Microbiol* **2011**, *77* (17), 6148-57.
123. Fernández-Arrojo, L.; Marín, D.; Gómez De Segura, A.; Linde, D.; Alcalde, M.; Gutiérrez-Alonso, P.; Ghazi, I.; Plou, F. J.; Fernández-Lobato, M.; Ballesteros, A., Transformation of maltose into prebiotic isomaltooligosaccharides by a novel α -glucosidase from *Xanthophyllomyces dendrorhous*. *Process Biochemistry* **2007**, *42* (11), 1530-1536.
124. Bergfors, T., *Protein Crystallization*. Second Edition ed.; International University Line: La Jolla, California, **2009**; p 500.
125. Rupp, B., *Biomolecular Crystallography. Principles, Practice, and Application to Structural Biology*. 1st Edition ed.; Garland Science: **2009**; p 800.
126. McPherson, A., Introduction to protein crystallization. *Methods* **2004**, *34* (3), 254-265.
127. Bergfors, T., Seeds to crystals. *J Struct Biol* **2003**, *142* (1), 66-76.
128. Wernimont, A.; Edwards, A., In situ proteolysis to generate crystals for structure determination: an update. *PLoS One* **2009**, *4* (4), e5094.
129. Stura, E. A.; Nemerow, G. R.; Wilson, I. A., Strategies in the crystallization of glycoproteins and protein complexes. *Journal of Crystal Growth* **1992**, *122* (1), 273-285.
130. Martinez-Ripoll, M. Crystallography-Cristalografía. <http://www.xtal.iqfr.csic.es/Cristalografía>.
131. Hassell, A. M.; An, G.; Bledsoe, R. K.; Bynum, J. M.; Carter, H. L., 3rd; Deng, S. J.; Gampe, R. T.; Grisard, T. E.; Madauss, K. P.; Nolte, R. T.; Rocque, W. J.; Wang, L.; Weaver, K. L.; Williams, S. P.; Wisely, G. B.; Xu, R.; Shewchuk, L. M., Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **2007**, *63* (Pt 1), 72-9.
132. Matthews, B. W., Solvent content of protein crystals. *J Mol Biol* **1968**, *33* (2), 491-7.
133. Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A., X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **2007**, *40* (3), 191-285.

134. Rambo, R. P.; Tainer, J. A., Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu Rev Biophys* **2013**, *42*, 415-41.
135. Svergun, D.; Barberato, C.; Koch, M. H. J., CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **1995**, *28* (6), 768-773.
136. Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D.; Konarev, P. V.; Svergun, D. I., New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **2012**, *45* (Pt 2), 342-350.
137. Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Gallo Cassarino, T.; Bertoni, M.; Bordoli, L.; Schwede, T., SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **2014**, *42* (Web Server issue), W252-8.
138. Nielsen, M.; Lundegaard, C.; Lund, O.; Petersen, T. N., CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res* **2010**, *38* (Web Server issue), W576-81.
139. Bowie, J. U.; Luthy, R.; Eisenberg, D., A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253* (5016), 164-70.

9. ANEXOS

ANEXO I: RESUMEN DE LOS OLIGOSACÁRIDOS REFERIDOS EN LA TESIS

En la siguiente tabla se encuentra un resumen de los mono-, di-, tri-, oligo- y polisacáridos contenidos en esta tesis, indicando su nombre común y su composición química.

	Nombre común	Nombre sistemático
Monosacáridos	Fructosa	D-fructofuranosa
	Galactosa	D-galactopiranososa
	Glucosa	D-glucopiranososa
	Manitol	D-manitol (monosacárido lineal)
	Sorbitol	D-glucitol (monosacárido lineal)
	Xilosa	D-xilopiranososa
Disacáridos	Celobiosa	β -D-glucopiranosil-(1-4)-D-glucopiranososa
	Gentiobiosa	β -D-glucopiranosil-(1-6)-D-glucopiranososa
	Gentiobiulosa	β -D-glucopiranosil-(1-6)-D-fructofuranosa
	Isomaltosa	α -D-glucopiranosil-(1-6)-D-glucopiranososa
	Isomaltulosa (Palatinosa)	α -D-glucopiranosil-(1-6)-D-fructofuranosa
	Kojibiosa	α -D-glucopiranosil-(1-2)-D-glucopiranososa
	Lactosa	α -D-galactopiranosil-(1-4)-D-glucopiranososa
	Lactulosa	β -D-galactopiranosil-(1-4)-D-fructofuranosa
	Laminaribiosa	β -D-glucopiranosil-(1-3)-D-glucopiranososa
	Maltosa	α -D-glucopiranosil-(1-4)-D-glucopiranososa
	Maltulosa	α -D-glucopiranosil-(1-4)-D-glucopiranososa
	Melibiosa	α -D-galactopiranosil-(1-6)-D-glucopiranososa
	Melibiulosa	α -D-galactopiranosil-(1-6)-D-fructofuranosa
	Nigerosa (Sakebiosa)	α -D-glucopiranosil-(1-4)-D-fructofuranosa
	Sacarosa	β -D-fructofuranosil-(2-1)- α -D-glucopiranososa
	Soforosa	β -D-glucopiranosil-(1-2)-D-glucopiranososa
	Trehalosa	α -D-glucopiranosil-(1-1)- α -D-glucopiranososa
	Turanosa	β -D-fructofuranosil-(3-1)- α -D-glucopiranososa
	Xilobiosa	β -D-xilopiranosil-(1-4)-D-xilopiranososa
Trisacáridos	Erlosa	β -D-fructofuranosil-(2-1)- α -D-glucopiranosil-(4-1)- α -D-glucopiranososa
	Isomaltotriosa	α -D-glucopiranosil-(1-6)- α -D-glucopiranosil-(1-6)-D-glucopiranososa
	1-Kestosa	β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- α -D-glucopiranososa
	6-Kestosa	β -D-fructofuranosil-(2-6)- β -D-fructofuranosil-(2-1)- α -D-glucopiranososa
	Lactosacarosa	β -D-fructofuranosil-(2-1)- α -D-glucopiranosil-(4-1)- β -D-galactopiranososa

	Maltotriosa	α -D-glucopiranosil-(1-4)- α -D-glucopiranosil-(1-4)-D-glucopiranososa
	Neokestosa	β -D-fructofuranosil-(2-6)- α -D-glucopiranosil-(1-2)- β -D-fructofuranosa
	Neoerlosa	β -D-fructofuranosil-(2-6)- α -D-glucopiranosil-(4-1)- α -D-glucopiranososa
	Panosa	α -D-glucopiranosil-(1-6)- α -D-glucopiranosil-(1-4)-D-glucopiranososa
	Rafinosa	β -D-fructofuranosil-(2-1)- α -D-glucopiranosil-(6-1)- α -D-galactopiranososa
Oligosacáridos (>3 unidades)	Estaquiosa	β -D-fructofuranosil-(2-1)- α -D-glucopiranosil-(6-1)- α -D-galactopiranosil-(6-1)- α -D-galactopiranososa
	Fructosil-nistosa	β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- α -D-glucopiranososa
	Maltotetraosa	α -D-glucopiranosil-(1-4)- α -D-glucopiranosil-(1-4)- α -D-glucopiranosil-(1-4)-D-glucopiranososa
	Neonistosa	β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- α -D-glucopiranosil-(6-2)- β -D-fructofuranosa
	Nistosa	β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- β -D-fructofuranosil-(2-1)- α -D-glucopiranososa
Polisacáridos (Polímeros) siendo n el número de repeticiones	Amilosa	$[\alpha$ -D-glucopiranosil-(1-4)- α -D-glucopiranosil] $_n$
	Amilopectina	$[\alpha$ -D-glucopiranosil-(1-4)- α -D-glucopiranosil] $_n$, con ramificaciones α -(1-6)
	Celulosa	$[\beta$ -D-glucopiranosil-(1-4)- β -D-glucopiranosil] $_n$
	Dextrano	$[\alpha$ -D-glucopiranosil-(1-6)- α -D-glucopiranosil] $_n$, con ramificaciones α -(1-4)
	β -Glucano	$[\beta$ -D-glucopiranosil-(1-4)/(1-3)- β -D-glucopiranosil] $_n$
	Inulina	$[\beta$ -D-fructofuranosil-(2-1)- β -D-fructofuranosil] $_n$ -(2-1)- α -D-glucopiranososa
	Levano	$[\beta$ -D-fructofuranosil-(2-6)- β -D-fructofuranosil] $_n$ -(2-6)-D-glucopiranososa

ANEXO II: CRISTALIZACIÓN Y RESOLUCIÓN DE PROTEÍNAS

La Biología Estructural es un ámbito de la ciencia cuyo objetivo es estudiar la estructura de macromoléculas biológicas, tales como los ácidos nucleicos o las proteínas, y su relación con la función biológica. Esta área científica utiliza técnicas de la Biología Molecular, la Bioquímica y la Biofísica para determinar la forma tridimensional de las macromoléculas biológicas. En el marco de esta tesis, la resolución estructural de proteínas se realiza mediante Cristalografía de rayos X, una de las técnicas experimentales más potentes para determinar la estructura tridimensional de las macromoléculas biológicas a nivel atómico y sin límite en el tamaño molecular. Esta técnica consta de varias etapas, resumidas en la figura A.1.

En primer lugar, se utilizan los métodos de la Biología Molecular para aislar y clonar el gen de interés y, posteriormente, expresar y purificar la proteína de estudio. Para que una muestra de proteína sea adecuada para su cristalización es necesario que presente un alto grado de pureza (90-95%), que sea estable, homogénea y se pueda obtener en cantidades suficientes¹²⁴. Cada proteína seguirá un proceso exclusivo de obtención en base a sus propiedades físico-químicas y biológicas, lo que supone que, en determinadas ocasiones, esta etapa actúe como “cuello de botella” del proceso. Para facilitar esta cuestión, un análisis bioinformático previo puede proporcionar información sobre parámetros que hay que controlar en la purificación y cristalización de una proteína, como la solubilidad, el correcto plegamiento y las regiones de desorden, así como obtener información estructural proveniente de proteínas homólogas.

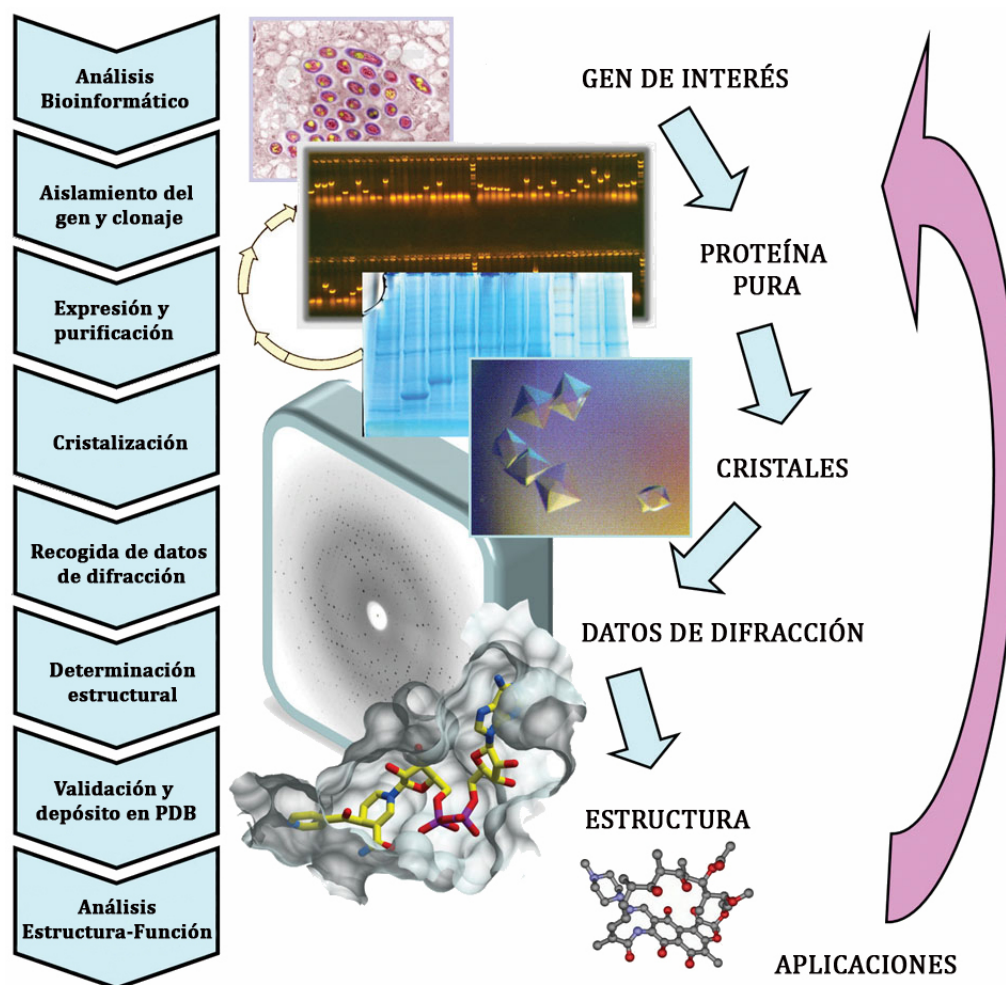


Figura A.1. Etapas básicas de la resolución estructural de proteínas mediante Cristalografía de rayos X. Imagen adaptada de ¹²⁵.

CRISTALIZACIÓN

Tras la etapa de purificación, la proteína se encuentra disuelta en una solución tamponadora (20-50 mM), a un pH compatible con la proteína, y suplementada con sales como el cloruro de sodio (50-150 mM) para mantener la fuerza iónica del medio y aumentar así la solubilidad y estabilidad de la proteína. Si se han cumplido las premisas de pureza, estabilidad y homogeneidad, la probabilidad de conseguir cristales se incrementa. Como la proteína es termodinámicamente estable en esta solución, se mantendrá soluble siempre y cuando su concentración no alcance el punto de saturación (curva de solubilidad). Cuando dicho punto se sobrepasa, la solución está sobresaturada, siendo termodinámicamente metaestable. En este momento se inicia un proceso de agregación molecular que, si se controla, llevará a la obtención de cristales (nucleación). En cambio, si la agregación no se controla, se llega a una situación termodinámicamente inestable y se produce una precipitación masiva que da lugar a un precipitado amorfo, lejos del orden inherente a un cristal¹²⁵ (figura A.2).

Para producir la sobresaturación de la solución, se utilizan distintos agentes precipitantes, que ayudan a la agregación progresiva de la proteína. Lo más habitual son los polímeros de diferente peso molecular, como el polietilenglicol (PEG), o las sales a alta concentración. También suelen utilizarse alcoholes, que modifican la constante dieléctrica del solvente, o compuestos volátiles, cuya evaporación altera la solubilidad de la proteína paulatinamente.

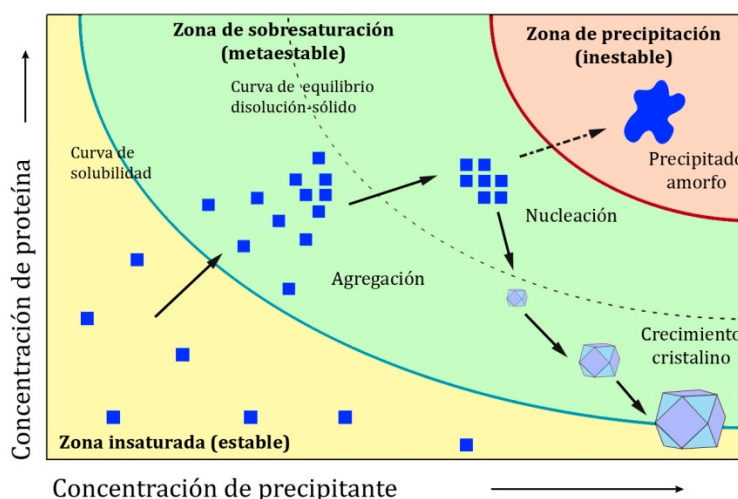


Figura A.2. Diagrama de fases del proceso de cristalización. Las flechas continuas indican el proceso que siguen las moléculas hasta formar los cristales.

Existen diversas técnicas que permiten llevar a cabo experimentos controlados de cristalización de macromoléculas. En esta tesis, la técnica empleada en todos los casos ha sido la difusión de vapor, donde un pequeño volumen de la solución de proteína se pone en contacto con una cantidad similar de una solución de cristalización que contiene los agentes precipitantes, y se expone a un ambiente que contiene un volumen mayor de dicha solución de cristalización (reservorio), sin que la gota y el reservorio entren en contacto, y todo ello aislado del exterior¹²⁵ (figura A.3). En la gota, el agente precipitante se encuentra diluido frente a la concentración que presenta en el reservorio. Bajo estas condiciones, se producirá una evaporación de agua desde la gota hacia el reservorio, hasta equilibrar la concentración del agente precipitante en ambas soluciones. Así pues, el volumen de la gota irá disminuyendo, mientras que la proteína se va concentrando. Por tanto, el sistema se acerca lentamente a la saturación y, posteriormente, a la sobresaturación, pasando por la zona de nucleación en la que se generarán los primeros núcleos, *clusters* de moléculas de proteína agregadas de forma ordenada. Este fenómeno hace que disminuya la concentración de la proteína soluble, llevando el sistema a una situación en la que el crecimiento de los núcleos requiere menos energía

que la formación de núcleos nuevos. Así pues, la nucleación se detiene y se inicia la etapa de crecimiento del cristal, durante la cual la proteína soluble se une a la superficie de los núcleos preexistentes, aumentando su tamaño. Finalmente, el sistema se equilibra al llegar a la curva de solubilidad, deteniéndose el crecimiento del cristal¹²⁶ (figura A.2).

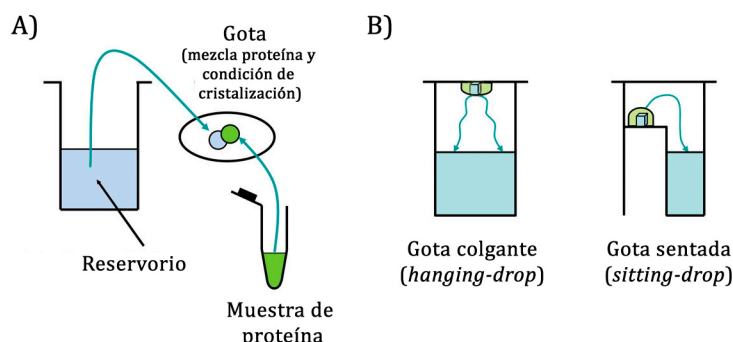


Figura A.3. Experimento de cristalización por difusión de vapor. A) Mezcla de un pequeño volumen de muestra de proteína (verde) y solución de cristalización (azul) para el ensayo. B) Configuración del experimento de cristalización en gota colgante y en gota sentada. Imagen adaptada de ¹²⁵.

No se pueden conocer las condiciones de cristalización en las que se van a obtener cristales *a priori*. Por ello, en una primera etapa, se hace un cribado con técnicas de alto rendimiento (HTS, *High-throughput screening*) de numerosas soluciones de cristalización utilizando un robot (en nuestro caso, Innovadine NanodropTM ExtY) y diversos *kits* comerciales (descritos en las distintas publicaciones de esta tesis), aplicando la técnica de difusión de vapor con configuración de gota sentada sobre placas de 96 pocillos InnovaplateTM SD-2, que permite usar volúmenes de gota muy pequeños (250-500 nL).

Generalmente, los cristales que se obtienen en esta fase son pequeños y suelen presentar defectos e irregularidades. Por ello, tras la búsqueda de condiciones de cristalización iniciales, se realiza una etapa de optimización que, dependiendo de cada caso particular, puede conllevar más o menos pasos. Una de las técnicas de optimización más habituales es el escalado, que consiste en aumentar los volúmenes de proteína y precipitante tanto en la gota (hasta 1 o 2 μ L) como del reservorio, con el objetivo de conseguir cristales de mayor tamaño que faciliten la toma de datos de difracción. Además, para controlar la cristalización y el crecimiento cristalino es fundamental manipular ciertas variables del experimento, que incluyen la concentración de la proteína, el pH, la temperatura y el agente precipitante (tanto el tipo como su concentración). A veces, se añaden otras moléculas, denominadas aditivos, que estabilizan las proteínas y favorecen su empaquetamiento. Estos aditivos incluyen alcoholes, detergentes, aminoácidos y azúcares, entre otros¹²⁶.

Asimismo, existe una técnica denominada sembrado (*seeding*) que consiste en introducir semillas de cristales en una gota pre-incubada de proteína en las condiciones de cristalización, que sirven como base para el crecimiento de nuevos cristales. Suele mejorar notablemente la calidad de los mismos¹²⁷, y reducir el tiempo de crecimiento. Estas semillas se consiguen triturando cristales preformados que no son suficientemente buenos para tomar datos de difracción, pero suficientes para introducir núcleos en nuevos experimentos de cristalización. De esta forma se puede realizar un proceso de crecimiento controlado sobre núcleos preformados.

Por último, se han descrito otras técnicas que facilitan la cristalización de proteínas que presentan problemas en la obtención de cristales o en su reproducibilidad. La proteólisis *in situ*¹²⁸, por ejemplo, consiste en introducir proteasas en las gotas de los experimentos de cristalización, con el fin de eliminar bucles expuestos o que interfieran en los contactos cristalinos necesarios para formar el cristal en el proceso de cristalización.

Por otra parte, cuando las proteínas son expresadas en sistemas eucariotas, la presencia de glicosilaciones puede disminuir la calidad de los cristales ya que los azúcares presentan mucha heterogeneidad química y estructural. Por ello, para la eficiente cristalización de las glicoproteínas, se procede a su desglicosilación previa¹²⁹.

PREPARACIÓN DE LAS MUESTRAS Y RECOGIDA DE DATOS DE DIFRACCIÓN

Una vez que se obtiene un cristal de buena calidad, se procede a la toma de datos de difracción y a su resolución estructural. Para ello debemos recurrir al fenómeno de difracción que se produce al incidir un haz de rayos X sobre el cristal, generando un patrón de haces difractados a distintos ángulos y con diferentes intensidades.

La microscopía de luz visible es incapaz de proporcionar información a la escala de las estructuras atómicas de las moléculas que componen las proteínas. Es por ello que la cristalografía de rayos X es la técnica indicada, ya que se trata de una radiación electromagnética con una longitud de onda en el rango de 0,1 – 100 Å. En los experimentos de difracción de rayos X para la resolución de estructuras de macromoléculas se utiliza una banda que abarca, aproximadamente, desde los 0,6 Å hasta los 2,5 Å. La elección de este rango se debe a que equivale a la magnitud de los enlaces existentes entre los átomos de las moléculas que conforman el cristal, por lo que la interacción de esta radiación con la materia es capaz de devolver información a resolución atómica¹²⁵. La obtención de rayos X con estas características puede llevarse a cabo en generadores de ánodo rotatorio de laboratorio o en las instalaciones de radiación sincrotrón (figura A.4). Las medidas de difracción tomadas a lo largo de este trabajo de tesis se han llevado a cabo en los sincrotrones ESRF (Grenoble, Francia), Petra-DESY (Hamburgo, Alemania), Diamond (Didcot, Reino Unido) y ALBA (Barcelona, España).

La señal que proporciona cada molécula, cada proteína de forma aislada, es muy pequeña. Por ello, al utilizar cristales se consigue un efecto multiplicador que aumenta la señal de difracción, ya que las macromoléculas se encuentran todas con la misma orientación. La preparación de las muestras cristalinas requiere de tres pasos: crio-protección y montaje del cristal en el soporte, enfriamiento rápido del cristal (*shock-cooling*) y trasvase del cristal a la cabeza goniométrica del equipo de rayos X (figura A.4). Los cristales de proteína se recogen de la placa de cristalización con un lazo de nailon de tamaño similar al de cada cristal en particular (0,025-0,5 µm), unido a un soporte magnético. Para reducir el daño que sufre el cristal por la radiación incidente, los cristales son vitrificados a una temperatura de 100 K. Debido al alto contenido en agua de los cristales de proteína, estos se tratan con una solución crioprotectora antes de sumergirlos en nitrógeno líquido. El crioprotector actúa desplazando el agua no estructural del cristal y evitando así la formación de hielo. Los agentes crioprotectores más usados son: PEG, MPD (2-metil-2,4-pentanodiol), glicerol, paratona o criosales. Finalmente, el soporte se monta sobre un goniómetro que permitirá su centrado y rotación durante la recogida de los datos.

Los rayos X difractados por el cristal se registran utilizando un detector. Existen distintos detectores, pero el más utilizado actualmente en los sincrotrones es el detector de matriz de píxeles, que contiene una capa de silicio muy sensible que absorbe los rayos X y los convierte en una señal eléctrica. Además, el sistema de lectura es extremadamente eficiente y rápido, y genera muy poco ruido de fondo.

Cuando un haz de rayos X incide sobre un cristal, interacciona con los electrones de los átomos que componen el cristal, haciéndolos vibrar acopladamente con las variaciones periódicas de su campo eléctrico. Al oscilar, los electrones se convierten en focos de nueva radiación X que se emite de forma esférica, fenómeno denominado dispersión, normalmente elástica. Los rayos X dispersados

por los electrones interfieren entre sí, pudiendo dar lugar a anulaciones. Sin embargo, como los átomos en un cristal están ordenados de manera regular y periódica, también ocurre que, en determinadas direcciones, se combinan y refuerzan para dar lugar a lo que se denomina dispersión cooperativa o difracción.

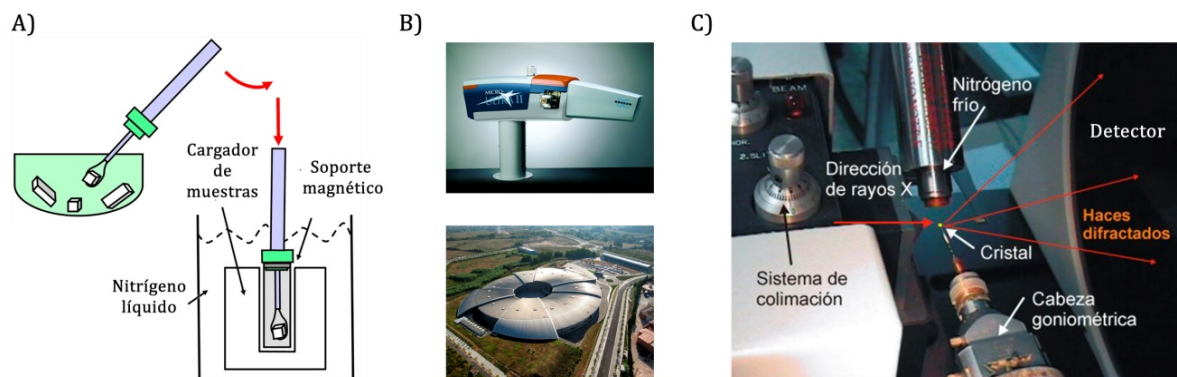


Figura A.4. Preparación de la muestra cristalizada para su difracción. (A) Montaje del cristal en el soporte magnético e inmersión en nitrógeno líquido para su conservación. Imagen adaptada de ¹²⁵. (B) Fuente de rayos X: ánodo rotatorio (arriba) y radiación sincrotrón (abajo). (C) Detalle del montaje del cristal para el experimento de difracción. Imágenes adaptadas de ¹³⁰.

CONCEPTO DE CRISTAL, CELDILLA UNIDAD Y UNIDAD ASIMÉTRICA.

Debido a que las moléculas de proteína de un cristal se encuentran ordenadas de forma periódica, es posible determinar una zona mínima del cristal que se repite por traslación en las tres direcciones espaciales. Se entiende por “cristal” a una red imaginaria formada por una unidad primaria de volumen, denominada “celdilla unidad” o “celdilla elemental”, que se repite periódicamente por traslación a lo largo de las tres direcciones del espacio. Esta celdilla queda definida por tres ejes (a , b , c) y por tres ángulos (α , β , γ), y sirve como sistema de coordenadas para referir las posiciones de cada átomo de la estructura. En el interior de la celdilla elemental pueden existir, a su vez, operaciones de repetición que se denominan “elementos de simetría”, como ejes de rotación, planos de reflexión y planos de deslizamiento. Así, la “unidad asimétrica” es el conjunto de moléculas mínimo que, tras aplicar los elementos de simetría, representa el contenido de la celdilla unidad. Las formas en que se pueden repetir esta unidad asimétrica no son infinitas y las macromoléculas biológicas sólo presentan 65 modalidades de repetición distintas, denominadas “grupos espaciales” (figura A.5).

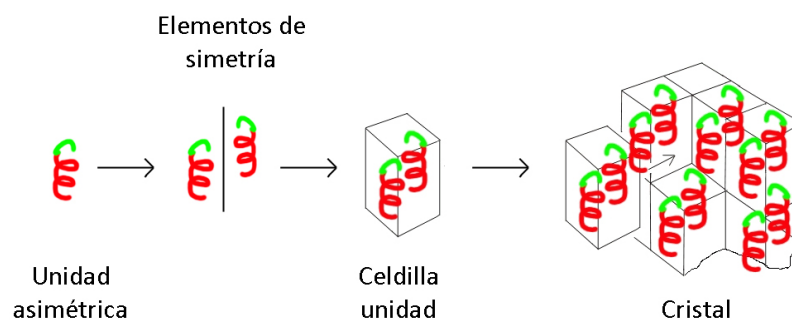


Figura A.5. Concepto de unidad asimétrica, celdilla unidad y cristal. Imagen adaptada de ¹³⁰.

RESOLUCIÓN ESTRUCTURAL DE LA PROTEÍNA

Una vez que se ha obtenido el patrón de difracción experimental, se suceden varias etapas para analizar la información y extraer la estructura tridimensional de la proteína, resumidas en la figura A.6.

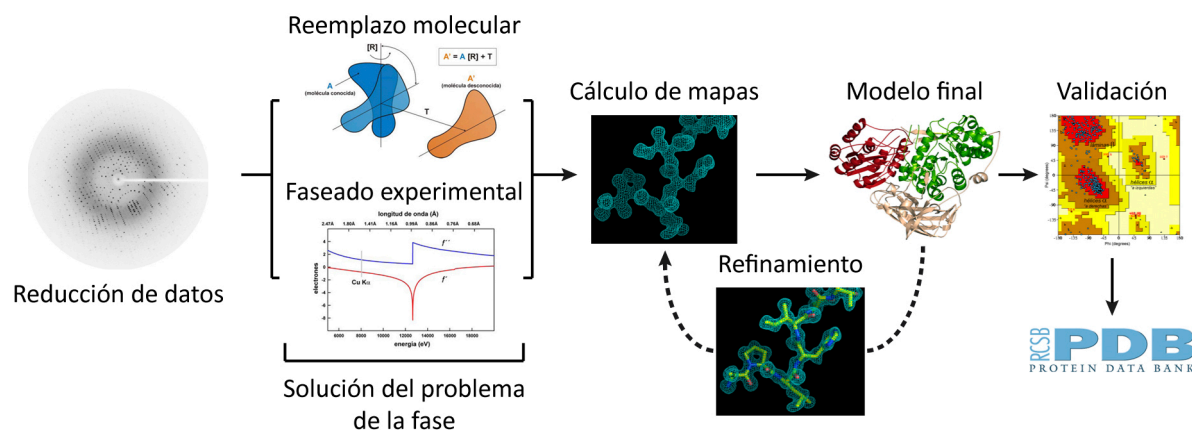


Figura A.6. Etapas del procesamiento de un patrón de difracción.

En primer lugar, se utilizan algunas imágenes del patrón de difracción para extrapolar la geometría de la red cristalina a partir de las posiciones de las reflexiones sobre el patrón, proceso que se conoce como “indexado”. De esta manera se calculan las dimensiones de la celdilla unidad (a , b , c , α , β , γ) y se estima el posible grupo espacial característico del cristal. Hay que tener en cuenta que, al registrar una rotación de un cristal que contiene diferentes elementos de simetría, cada reflexión puede medirse varias veces en distintos momentos del experimento. Esta redundancia es imprescindible para un correcto procesamiento de los datos. Tras el indexado se realiza una etapa de “integrado” consistente en estimar la intensidad (I_{hkl}) para cada reflexión con posición (hkl) a lo largo de todas las imágenes registradas, agrupando la información de las múltiples veces que ha sido registrada a lo largo de la rotación del cristal. Debido a posibles defectos de heterogeneidad del cristal o al daño por radiación, no se puede considerar que los datos tomados al inicio del experimento o en determinada zona del cristal vayan a presentar la misma calidad que los tomados al final del experimento o en otra zona de incidencia del haz durante la rotación. Por tanto, debido a esta falta de correlación, se hace necesaria una etapa de “escalado”.

Posteriormente, se debe solucionar el “problema de las fases”, ya que de los patrones de difracción sólo se extraen intensidades absolutas por lo que no se puede calcular directamente la función de densidad electrónica que precisa del ángulo o fase asociado a cada reflexión. Así, cada reflexión detectada en el patrón de difracción tiene asociado un factor de estructura F_{hkl} derivado de la dispersión del haz incidente por determinados átomos del cristal. Por tratarse de una onda, el factor de estructura F_{hkl} es una magnitud vectorial que se compone de módulo ($|F_{hkl}|$) y fase (ϕ_{hkl}). El módulo del factor de estructura de cada reflexión es proporcional a su intensidad, por lo que el valor de $|F_{hkl}|$ se extrae directamente del análisis del patrón de difracción tras la etapa de escalado. En cambio, durante la recogida de datos se pierde la información de su fase, dando lugar al denominado “problema de la fase”. Por tanto, es necesario utilizar técnicas específicas que nos estimen unas fases iniciales aproximadas, necesarias para la construcción de un modelo inicial de nuestra proteína, que serán mejoradas progresivamente hasta obtener un modelo final que se ajuste lo más posible al experimento. Existen esencialmente dos formas de calcular la fase, el método del “reemplazo molecular”, utilizado en esta tesis, y el “faseado experimental”¹²⁵.

El reemplazo molecular consiste en utilizar como fases iniciales las obtenidas de otra proteína similar a la que se pretende resolver, presuponiendo que secuencias similares se organizan en

plegamientos comparables. Para ello, el modelo tridimensional de la proteína similar (modelo de reemplazo) se enfrenta con los datos experimentales del cristal de la proteína a resolver (escalado) y, mediante el cálculo de funciones de rotación y traslación, se posiciona en la red cristalina a resolver. Cuando no se dispone de una proteína previamente resuelta similar a la que queremos resolver, se recurre al faseado experimental, donde se introducen átomos en el cristal que modifican el patrón de difracción de tal forma que se pueda calcular primero una subestructura formada por dichos átomos y, posteriormente, tomando esta subestructura como referencia, construir la estructura de la proteína.

Una vez asignada la fase a cada reflexión, se puede calcular la función de densidad electrónica del cristal. La representación gráfica de esta función origina el mapa de densidad electrónica ($2F_o - F_c$), el cual proporciona la información necesaria para la construcción del modelo de la proteína, así como el mapa de densidad electrónica diferencia ($F_o - F_c$), donde los picos de densidad positiva revelan que el modelo está incompleto, mientras que los negativos muestran posiciones incorrectas del mismo.

Posteriormente, se realizan varios ciclos de refinamiento, hasta obtener un modelo estructural final que explique el espectro de difracción con la mayor precisión posible. El refinamiento es una etapa imprescindible que consiste en minimizar las diferencias entre los factores de estructura calculados del modelo con los factores de estructura experimentales, para obtener un nuevo mapa de densidad electrónica de mejor calidad que el anterior que permite construir un nuevo modelo mejorado. Además, para mejorar las fases, se utilizan dos técnicas cristalográficas muy eficaces: el aplanamiento del solvente y la simetría no cristalográfica (NCS). El aplanamiento del solvente genera una máscara que distingue proteína y solvente, fijando un valor constante para la densidad del solvente, lo que elimina ruido en el mapa de densidad. Por su parte, el promediado de densidad utilizando la simetría no cristalográfica se aplica cuando existen varias copias de la molécula de proteína dentro de la unidad asimétrica, de manera que se usa esta redundancia en la información para promediar y minimizar errores en el cálculo del mapa.

Una vez resuelta la estructura, y antes de ser depositada en la base de datos PDB (*Protein Data Bank*), debe pasar un minucioso proceso de validación donde se analiza la geometría del modelo estructural final y su adecuación al espectro de difracción experimental.

CRISTALIZACIÓN DE COMPLEJOS PROTEÍNA-LIGANDO

Para poder cristalizar la proteína en complejo con sus substratos, es necesario inactivar la enzima, por lo que hay que retomar la etapa de clonación y expresión para obtener el mutante inactivo de la enzima por mutagénesis dirigida, y a partir de ahí repetir los pasos de cristalización y resolución estructural (figura A.1). Existen dos aproximaciones para realizar la cristalización en presencia de ligandos: la inmersión (*soaking*) y la co-cristalización¹³¹. El *soaking* consiste en sumergir los cristales de proteína preformados en una solución que contiene las mismas condiciones de cristalización en que crecieron, suplementada con el ligando a introducir. La co-cristalización consiste en la incubación de la proteína con el ligando antes de realizar los experimentos de cristalización, que pueden conducir a nuevas condiciones distintas a las que originan cristales de la proteína libre. De este modo, los cristales obtenidos contienen el complejo formado por la proteína y el substrato. Los cristales de proteína contienen grandes cantidades de solvente (entre el 30 y el 80% del volumen del cristal)¹³². Así, estos canales de solvente permiten que el ligando difunda por el interior del cristal y pueda acceder al sitio de unión. Dependiendo de cada caso particular se utilizará una aproximación diferente, variando los tiempos de incubación o inmersión y la concentración de ligando.

ANEXO III: OTRAS TÉCNICAS DE BIOLOGÍA ESTRUCTURAL

La superación con éxito de la etapa de cristalización es una condición *sine qua non* para poder resolver su estructura por Cristalografía de rayos X. Es por eso que la obtención de buenos cristales y/o buenos patrones de difracción a veces se convierten en un paso limitante en el proceso de la resolución estructural de la proteína, y es necesario volver a la etapa de expresión y purificación para mejorar la muestra, los cristales y los patrones de difracción.

Por ello, en ocasiones, es necesario recurrir a otras técnicas dentro de la Biología Estructural que no requieran la cristalización de la proteína.

DIFRACCIÓN DE RAYOS X A BAJO ÁNGULO (SAXS)

La dispersión de rayos X a bajo ángulo (SAXS) es una técnica analítica que permite caracterizar el tamaño y la forma de una proteína. Frente a la cristalografía, presenta la ventaja de utilizar las proteínas en solución, evitando el a veces problemático proceso de cristalización. Sin embargo, la resolución estructural que de estos datos se puede obtener es mucho menor, llegando a resoluciones de unos 30 Å.

Aunque la cristalografía suministra un detalle estructural fundamental para análisis mecanicistas, describe conformaciones de baja energía de macromoléculas dentro de las redes cristalinas. Sin embargo, SAXS ofrece información del comportamiento en solución de la proteína: el plegamiento macromolecular, o el no-plegamiento, la agregación, las conformaciones extendidas, los dominios unidos por segmentos con flexibilidad, la forma, la conformación y el estado de oligomerización. Además, los experimentos de SAXS pueden llevarse a cabo mucho más rápidamente que cualquier otro de crio-microscopía electrónica, y la información derivada de los datos de SAXS puede ser complementaria a las estructuras de alta resolución¹³³.

Por ello, es una técnica que complementa a la cristalografía para aquellas proteínas o complejos macromoleculares que, por diversas causas, no se han conseguido cristalizar, o sólo se ha conseguido cristalizar una parte, o una conformación. Así, SAXS, en combinación con la cristalografía de proteínas y/o la resonancia magnética nuclear (RMN), puede ofrecer una visión más completa de ciertos sistemas biológicos complejos: SAXS aporta información esencial sobre el tamaño y forma de las partículas, mientras que la cristalografía o RMN aportan la estructura tridimensional a alta resolución de uno o varios de los componentes.

Esta robusta técnica permite caracterizar macromoléculas biológicas en solución, basada en analizar la dispersión de rayos X producida por una macromolécula al paso del haz, a ángulos muy próximos a cero, y que permite aportar información en el rango de 1 a 200 nanómetros. Actualmente, los experimentos de SAXS con moléculas biológicas se realizan en los sincrotrones, principalmente. Las fuentes de luz sincrotrón ofrecen un haz de rayos X monocromático, bien colimado, puntual e intenso, cualidades que lo hacen inmejorable para hacer medidas a bajos ángulos. Otra ventaja adicional es que se puede cambiar la longitud de onda con facilidad¹³⁴.

SAXS y la cristalografía de rayos X son técnicas fundamentalmente similares. El haz de rayos X incidente interactúa con los electrones de todos los átomos en la macromolécula, originando un “patrón de dispersión”, es decir, haces de rayos X con diferentes intensidades a diferentes ángulos de dispersión, que permite caracterizar la estructura de las partículas (figura A.7.A). Una diferencia fundamental entre la dispersión de la solución y la cristalografía de rayos X se encuentra en el nivel de organización de moléculas diana durante la recogida de datos. En SAXS, la señal que llega al detector es un promedio de las señales de todas las moléculas en solución, cada una en una determinada orientación. Por tanto, esta señal es continua y radialmente simétrica (isotrópica, igual en todas las direcciones del detector) (figura A.7.B). Por el contrario, en la cristalografía de rayos X las moléculas están altamente organizadas dentro de la red cristalina lo que produce interferencias que dan lugar a máximos de difracción discretos (patrón de difracción) debido a que la red cristalina funciona como un amplificador y se produce una convolución de las señales de las posiciones atómicas¹³³ (figura A.7.C).

El requisito fundamental para la generación de buenos datos de difracción para SAXS es la presencia de una muestra homogénea, tanto en composición como en estado de oligomerización, monodispersa y sin agregados. En general, el consumo de muestra de proteína para la toma de datos de SAXS es pequeña (20-50 μL a una concentración de 5-10 mg/mL).

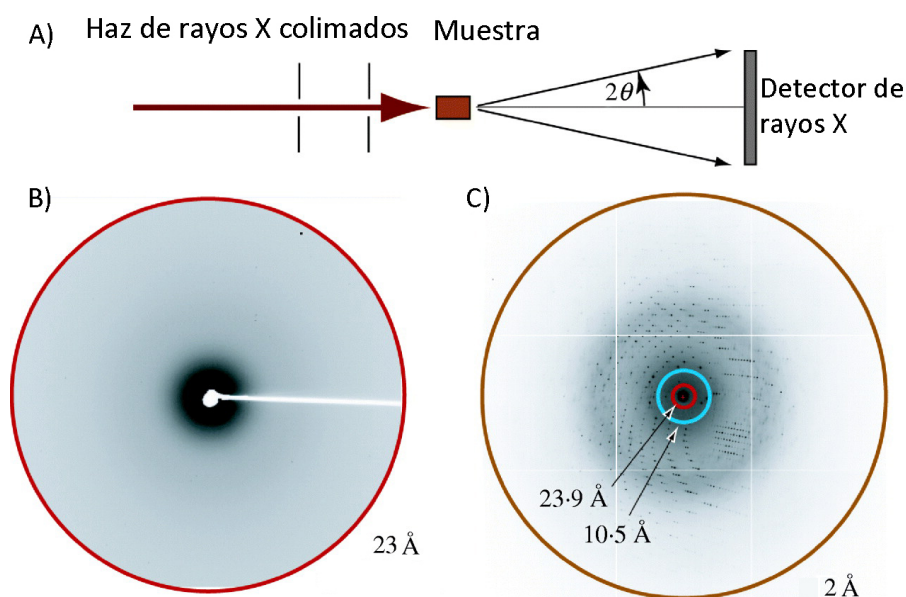


Figura A.7. Bases de la técnica de SAXS vs. Cristalografía de rayos X. A) Proceso de toma de datos de un experimento de SAXS o de Cristalografía de rayos X. Tanto las muestras de SAXS como los cristales (naranja) se colocan frente a un haz de rayos X colimado (rojo) y se miden los rayos X dispersados. El ángulo de cualquier dispersión, respecto del haz, es 2θ . B) Patrón de dispersión de una muestra en solución (SAXS), con una resolución de 23 Å. C) Patrón de difracción de un cristal, a 2 Å de resolución. En c) la resolución experimental que sería equivalente a la obtenida por SAXS se indica con un círculo rojo. El círculo azul indica la mayor resolución de SAXS alcanzable en SIBYLS (Lawrence Berkeley National Laboratories), 9 Å. Imagen adaptada de ¹³³.

PROCESADO Y ANÁLISIS DE LOS DATOS DE SAXS

En los experimentos de SAXS, la señal de dispersión detectada es una media de las densidades electrónicas de todos los átomos de la muestra (proteína + tampón). Por ello, a la media de dispersión de la muestra hay que restarle la media de dispersión del tampón (blanco). En cada experimento, se hacen varias medidas a distintas concentraciones. Luego, y tras restar el valor del blanco, dichas curvas se escalan para obtener una única curva de dispersión ideal para dicha muestra, con programas como PRIMUS, que también permiten analizar los datos obtenidos. Para ello, se utilizan, principalmente, la gráfica de Guinier, la Gráfica de Kratky y la función de distribución de pares, $P(r)$ ¹³³⁻¹³⁴ (figura A.8).

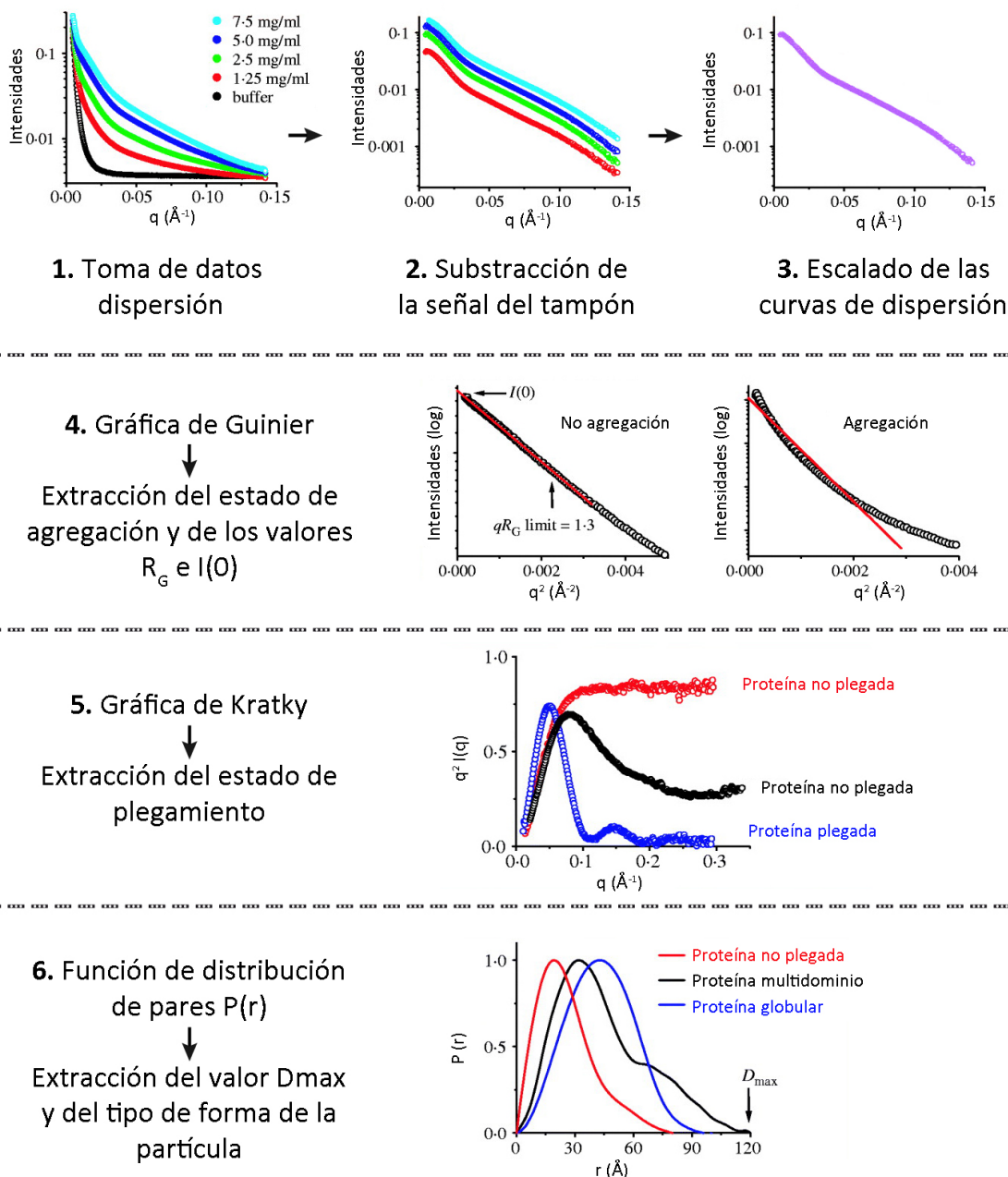


Figura A.8. Etapas en el procesamiento y análisis de los datos de SAXS.

La gráfica de Guinier aporta información sobre el estado de agregación de la muestra. La falta de linealidad es una señal de agregación. De esta representación lineal de los datos de dispersión se puede extraer los valores R_G e $I(0)$.

El radio de giro (R_G) es el radio promedio de cada partícula (molécula o complejo molecular), la distancia media de cada partícula desde el centro hasta cada extremo. Puesto que este valor es un promedio, no aporta información sobre la forma de la partícula y, por ello, no se puede usar para calcular el volumen ni la masa molecular de la partícula de interés.

$I(0)$ es la intensidad media a ángulo cero ($q=0$) y, calculado por extrapolación de la intensidad de dispersión a dilución infinita. Este valor está relacionado directamente con el volumen de la partícula y, por ello, se puede usar para calcular la masa molecular de dicha partícula.

La gráfica de Kratky, que también se calcula directamente de la curva de dispersión, es una buena herramienta para evaluar el plegamiento de las muestras. Para aquellas partículas bien plegadas, esta función resulta en un parábola. Las partículas desestructuradas presentan una gráfica más lineal, como ocurre en las muestras de proteínas desplegadas. Por ello, los datos de SAXS se pueden usar como análisis previos a la cristalización, para comprobar el correcto plegamiento de las proteínas.

La función de distribución de pares de densidades ($P(r)$, *pair-distribution function*) es una función de autocorrelación que se calcula directamente de la curva de dispersión mediante una transformada de Fourier, con programas como GNOM, y el resultado proporciona información directa sobre las distancias entre electrones de la partícula. Por ello, proporciona información sobre la forma de la partícula (globular o multidominio) y también si está completamente plegada.

De esta función $P(r)$ también se pueden calcular R_G e $I(0)$, con la ventaja de que $P(r)$ abarca todos los datos recolectados y no está limitado a una región más pequeña, como ocurre con la aproximación de Guinier. Si los valores de R_G e $I(0)$ calculados a partir de Guinier o de $P(r)$ coinciden, la muestra se comporta como partículas homogéneas y monodispersas. De la función $P(r)$ también se calcula el valor D_{max} , la máxima distancia lineal presente en la partícula.

Finalmente, el volumen de dicha partícula se puede calcular a partir de $I(0)$ (volumen de Porod) y, de este volumen, calcular la masa molecular de la partícula. Para proteínas globulares, el volumen de Porod (en nm^3) es el doble de la masa molecular (en kDa). Sin embargo, este cálculo no es adecuado para las partículas asimétricas.

MODELADO ESTRUCTURAL A PARTIR DE DATOS DE SAXS

La función $P(r)$ puede ser calculada para los modelos atómicos obtenidos por cristalografía de rayos X o RMN, y, aplicando la transformada de Fourier reversa, calcular una curva de dispersión de SAXS teórica. Esto es de gran utilidad para combinar los datos obtenidos de experimentos de SAXS con modelos atómicos cristalográficos.

Existen varios programas para calcular la curva de dispersión teórica de un modelo atómico y compararla con sus datos experimentales. CRY SOL¹³⁵ es uno de los programas más utilizados y utiliza el valor χ^2 como un indicador relativo de la bondad del ajuste entre la curva de dispersión teórica (*scattering* teórico) y los datos experimentales de SAXS de la proteína completa (*scattering* experimental). Valores bajos (idealmente <2) indican un buen ajuste.

Para crear modelos estructurales a partir de datos de SAXS de complejos moleculares donde se conocen los fragmentos que lo componen a nivel atómico, existen varios programas. En el caso de contener las partes, pero no tener información del segmento conector entre ellas, se puede utilizar el programa CORAL¹³⁶.

CORAL es un programa que genera un modelo macromolecular a partir de los fragmentos resueltos de dicho complejo y los datos experimentales de SAXS de la macromolécula completa. Para ello, utiliza una metodología de modelado por cuerpo rígido (*rigid-body modeling*) donde cada fragmento se considera un objeto estático y sólo se modifica su orientación espacial relativa. Primero, se fija uno de los fragmentos resueltos, normalmente, el de mayor tamaño. El resto de fragmentos se mueven alrededor de éste. Además, CORAL modela *ab initio* los segmentos conectores entre los fragmentos. En cada ciclo, se genera un modelo de la macromolécula y se evalúa su fiabilidad con CRY SOL. El movimiento de los fragmentos no-fijos está restringido por dos parámetros: 1) la longitud del enlace entre los fragmentos y 2) el valor de χ^2 de CRY SOL no puede exceder de 2. La selección del mejor modelo final se hace en base al menor valor de χ^2 alcanzado. Si bien un buen valor de χ^2 no garantiza una solución macromolecular correcta en este tipo de cálculos, como cada reconstrucción con CORAL parte de unas coordenadas y unas orientaciones distintas, la repetición en el resultado de distintos intentos indica que el modelo generado para la macromolécula se puede considerar correcto.

MÉTODOS DE PREDICCIÓN DE ESTRUCTURA TERCIARIA

Como ya se ha comentado anteriormente, los métodos de resolución de Biología Estructural requieren de muestras de proteína homogéneas. Cuando experimentalmente no es posible obtener la estructura tridimensional de una proteína, existen aproximaciones bioinformáticas que permiten modelar y predecir, con mayor o menor confianza, su estructura terciaria.

Por una parte, el modelado por homología (*homology modeling*), o modelado comparativo, de secuencia es la técnica más utilizada y presupone que aquellas secuencias que presentan un alto porcentaje de identidad también compartirán una estructura similar. Esta estrategia utiliza proteínas homólogas resueltas como modelo para predecir la estructura tridimensional de la proteína diana. Este es el caso de SWISS-MODEL¹³⁷, que funciona por alineamientos de secuencia con proteínas homólogas con altos porcentajes de identidad de secuencia. Sin embargo, cuando la secuencia de la proteína a modelar presenta menos de un 20% de identidad con las estructuras disponibles, el modelado por homología resulta muy complicado y menos eficiente. Por ello, algunos programas y servidores de modelado por homología han implementado la comparación de estructuras secundarias, como el servidor CPHmodel¹³⁸.

El modelado por subprocesos (*threading modeling*), o reconocimiento de plegamientos, se utiliza para aquellas proteínas que no presentan un homólogo estructural. En la naturaleza el número de plegamientos es finito (unos 1300 plegamientos distintos), por tanto, las proteínas a modelar deben contener alguno de esos plegamientos ya descritos. Primero, se predice la estructura secundaria de la proteína diana y luego, mediante análisis estadísticos y el alineamiento de las estructuras secundarias de la proteína diana con fragmentos de las estructuras depositadas, se identifican posibles plegamientos y se crea un modelo 3D. Sin embargo, este método puede resultar poco eficaz para proteínas que sólo contengan hebras β , ya que la predicción es menos fiable¹³⁹.